

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 07:48:47 ; Search time 1.76303 Seconds  
(without alignments)  
604.643 Million cell updates/sec

Title: US-09-857-308-5  
Perfect score: 37  
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	100.0	8	21 AAB03883	Human ART-1 peptid
2	37	100.0	8	22 AAB46941	Human ART-3 derlve
3	37	100.0	9	21 AAB03886	Human ART-1 peptid
4	37	100.0	363	22 AAG98646	Human cell death p
5	37	100.0	403	21 AAB45216	Gene 41 human secr
6	37	100.0	404	21 AAB45217	Human secreted pro
7	37	100.0	412	22 AAG98644	Human cell death p
8	37	100.0	414	21 AAB03880	Human ART-1 protei
9	37	100.0	419	22 AAM25845	Human protein sequ
10	32	86.5	100	23 ABB28637	Streptococcus poly

11	32	86.5	273	21	AA909399	Cenarchaeum symbio
12	31	83.8	269	21	AA909323	Cenarchaeum symbio
13	30	81.1	9	20	AA947894	Immunogenic peptid
14	30	81.1	9	22	AA22249	HIV peptide SEQ ID
15	30	81.1	9	22	AAU12541	Human HIV-1 Th-CTL
16	30	81.1	20	19	AAW55857	HIV p17 gag protei
17	30	81.1	20	22	AAU12335	Peptide fragment o
18	30	81.1	24	22	AAU12487	HIV Th-CTL peptide
19	30	81.1	24	23	AAU70250	HIV Th-CTL P17 epi
20	30	81.1	123	23	AAU10483	HIV Th-CTL epitope
21	30	81.1	139	22	ABB10879	Human pancreatic c
22	30	81.1	139	22	AA92721	Human digestive sy
23	30	81.1	247	22	AA922996	Human protein sequ
24	30	81.1	312	18	AAW20350	H. pylori outer me
25	30	81.1	312	18	AAW20857	H. pylori outer me
26	30	81.1	449	22	AA92706	Human protein sequ
27	30	81.1	577	22	AA93792	Human polypeptide,
28	30	81.1	756	22	AA087329	Novel central nerv
29	30	81.1	814	22	AA43565	Human polypeptide
30	29	78.4	9	20	AA47900	Immunogenic peptid
31	29	78.4	95	22	AAU14913	Novel bone marrow
32	29	78.4	158	22	AAU42587	Propionibacterium
33	29	78.4	269	17	AA88725	LIF3. Mycoplasma
34	29	78.4	560	20	AA99104	Human EAAT5 glutam
35	29	78.4	560	21	AA52196	Excitatory amino a
36	29	78.4	560	22	AAE10884	Human excitatory a
37	29	78.4	560	22	AAE11030	Human excitatory a
38	29	78.4	564	20	AA99105	Protein sequence S
39	29	78.4	564	21	AA52197	Excitatory amino a
40	29	78.4	564	22	AAE10885	Human excitatory a
41	29	78.4	572	17	AAW02256	Mouse L-glutamate
42	29	78.4	574	18	AAW26598	Human glutamate tr
43	29	78.4	574	19	AAW58553	Human excitatory a
44	29	78.4	574	20	AA42551	Human astrocyte wi
45	29	78.4	574	20	AA28296	Amino acid sequenc

#### ALIGNMENTS

RESULT 1  
AAB03883  
ID AAB03883 standard; protein; 8 AA.  
XX  
AC AAB03883;  
XX  
DT 26-OCT-2000 (first entry)  
XX  
DE Human ART-1 peptide fragment SEQ ID #5.  
XX  
DE Human; tumour antigen protein; ART-1; HLA antigen; cytostatic;  
KW bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.  
XX  
OS Homo sapiens.  
XX  
PN WO200032770-A1.  
XX  
PD 08-JUN-2000.  
XX  
PF 30-NOV-1999; 99WO-JP06682.  
XX  
PR 01-DEC-1998; 98JP-0341253.  
XX  
PA (SUMU) SUMITOMO PHARM CO LTD.  
XX  
PA (ITOH/) ITOH K.  
XX  
PI Itoh K, Gomi S;  
XX  
DR WPI; 2000-412318/35.  
XX  
PT Novel tumor antigen protein ART-1, tumor antigen peptide originating  
PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as  
PT remedies, preventives and diagnostics for tumors -

XX Claim 11; Page 35; 59pp; Japanese.

XX The invention relates to a novel human tumour antigen protein, ART-1.

CC Included in the invention are polynucleotide sequences encoding the ART-1

CC protein, and mutated ART-1 proteins which when broken down

CC intracellularly produce a tumour antigen peptide that can recognise HLA

CC antigen and bound cytotoxic T cells. Antibodies which specifically

CC recognise ART-1 and its derivative peptides, are also included in the

CC invention. ART-1 exhibits cytostatic activity. The tumour antigen

CC protein, tumour antigen peptide originating from it, their derivatives,

CC and DNAs are applicable in vivo or in vitro as remedies, preventives and

CC diagnostics for tumours.

CC The present sequence represents a fragment of the ART-1 protein, used in

CC the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 37; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 7.8e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 1 LYQAVATI 8

RESULT 2

AAB46941

ID AAB46941 standard; peptide; 8 AA.

XX AAB46941;

XX 04-MAY-2001 (first entry)

XX Human ART-3 derived tumor antigenic peptide SEQ ID 27.

XX Tumor antigenic peptide; interferon; IFN; antigen-specific T cell;

KW virucide; cytostatic; gene therapy; IFN-alpha; cytotoxic T cell;

KW viral infectious disease; ART-1; human.

XX Homo sapiens.

XX EP1074267-A1.

XX 07-FEB-2001.

XX 24-JUL-2000; 2000EP-0306263.

XX 22-JUL-1999; 99JP-0207687.

XX (SUMU ) SUMITOMO PHARM CO LTD.

XX Takasu H, Gotoh M, Yamaoka T;

XX WPI; 2001-193144/20.

XX Use of antigenic proteins, peptides, interferon or their encoding DNA,

PT in the manufacture of an agent for the induction of antigen-specific T

PT cells

XX Disclosure; Page 18; 25pp; English.

XX This invention describes the novel use of interferons (IFNs) or DNAs

CC capable of expressing the interferons and/or antigenic proteins (AP),

CC antigenic peptides derived from the proteins or DNAs capable of

CC expressing the antigenic proteins or peptides, in the manufacture of

CC an agent for induction of antigen-specific T cells. The products of

CC the invention have virucide and cytostatic activity and can be used for

CC gene therapy or as inducers of antigen-specific T cells. The action of

CC interferon-alpha (IFN-alpha) in a system for inducing specific cytotoxic

CC T cell (CTL) by administering an antigenic peptide in an incomplete

CC Freund's adjuvant (IFA) emulsion preparation form was evaluated. IFNs (or

CC DNA encoding IFNs) are useful in the manufacture of a medicament for

CC inducing antigen-specific T cells in an individual who has been

CC administered with AP (or DNA encoding AP) or vice versa. The medicament

CC is useful for the treatment or prophylaxis of a tumor or a viral

CC infectious disease.

XX Sequence 8 AA;

Query Match 100.0%; Score 37; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 7.8e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 1 LYQAVATI 8

RESULT 3

AAB03886

ID AAB03886 standard; protein; 9 AA.

XX AAB03886;

XX 26-OCT-2000 (first entry)

XX Human ART-1 peptide fragment SEQ ID #8.

XX Human; tumour antigen protein; ART-1; HLA antigen; cytostatic;

KW bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.

XX Homo sapiens.

XX WO200032770-A1.

XX 08-JUN-2000.

XX 30-NOV-1999; 99WO-JP06682.

XX 01-DEC-1998; 98JP-0341253.

XX (SUMU ) SUMITOMO PHARM CO LTD.

XX (ITOHI/) ITOH K.

XX Itoh K, Gomi S;

XX WPI; 2000-412318/35.

XX Novel tumor antigen protein ART-1, tumor antigen peptide originating

PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as

PT remedies, preventives and diagnostics for tumors

XX Claim 11; Page 51; 59pp; Japanese.

XX The invention relates to a novel human tumour antigen protein, ART-1.

CC Included in the invention are polynucleotide sequences encoding the ART-1

CC protein, and mutated ART-1 proteins which when broken down

CC intracellularly produce a tumour antigen peptide that can recognise HLA

CC antigen and bound cytotoxic T cells. Antibodies which specifically

CC recognise ART-1 and its derivative peptides, are also included in the

CC invention. ART-1 exhibits cytostatic activity. The tumour antigen

CC protein, tumour antigen peptide originating from it, their derivatives,

CC and DNAs are applicable in vivo or in vitro as remedies, preventives and

CC diagnostics for tumours.

CC The present sequence represents a fragment of the ART-1 protein, used in

CC the invention.

XX Sequence 9 AA;

Query Match 100.0%; Score 37; DB 21; Length 9;

Best Local Similarity 100.0%; Pred. No. 7.8e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

DB 1 LYQAVATI 8

RESULT 4

AAAG98646

ID AAC98646 standard; Protein: 363 AA.

XX AAC98646;

AC

XX 21-SEP-2001 (first entry)

DT

XX Human cell death protective cDNA clone CNI-00714 ORF3 protein, SEQ.81.

DE

XX Cell death protective; apoptosis; necrosis; human; drug screening;

XX cell death-associated disorder; central nervous system disorder;

KW psychiatric disorder; neurological disorder; ischaemia-related disorder;

KW stroke; cerebral infarction; ischaemic encephalopathy;

KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;

KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;

KW vascular disease; ophthalmological disorder; diabetic retinopathy;

KW macular degeneration; hypertension; myocardial infarction;

KW atherosclerosis; respiratory disorder; asthma; transgenic animal;

KW chronic obstructive pulmonary disease; neoplastic condition; cancer;

KW benign tumour; anaemia; gastrointestinal disorder; gastritis;

KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;

KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;

KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;

KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.

XX Homo sapiens.

OS

XX WO200145638-A2.

PN

XX 28-JUN-2001.

PD

XX 11-DEC-2000; 2000WO-US33547.

XX

XX 14-DEC-1999; 99US-0461697.

PR

XX (COGE-) COGENT NEUROSCIENCE INC.

XX

PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;

XX WPI; 2001-390297/41.

DR

XX N-PSDB; AAH84170, AAH84173.

XX

XX Novel protective sequence polynucleotides and polypeptides, used to

PT identify modulators of their expression and activity, which are used in

PT to treat central nervous system conditions, diseases and disorders -

XX

XX Claim 1; Fig 6C; 325pp; English.

PS

XX Sequences AAH84132-AAH84370 represent human nucleic acid sequences which

CC protect against cell death (i.e., apoptosis or necrosis). Sequences

CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,

CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,

CC while the remaining nucleic acid sequences within the range given above

CC represent the open reading frames (ORFs) of these cDNA clones. Sequences

CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death

CC protective ORFs. The cell death protective cDNA clones are able to

CC prevent, delay or reverse progression through the apoptotic or necrotic

CC pathways when injected into a cell predisposed to or undergoing cell

CC death. The cell death protective nucleic acids and polypeptides can be

CC used in the diagnosis and treatment of disorders associated with cell

CC death, and to screen for compounds which modulate their activity or

CC expression. Such modulators, preferably a small organic molecule, an

CC antibody, a ribozyme, or an antisense molecule, can also be used to treat

CC cell death-related diseases. Such diseases include those associated with

CC the central nervous system including psychiatric or neurological

CC disorders, especially ischaemia-related conditions such as strokes, and

CC also includes neurodegenerative disorders such as Alzheimer's disease,

CC Huntington's disease, or Parkinson's disease. The modulators may also be

CC

CC used to treat infections such as meningitis, malaria, or trypanosomiasis;

CC vascular diseases such as ischaemic encephalopathy or cerebral

CC infarction; eye conditions such as diabetic retinopathy or macular

CC degeneration; hypertension; myocardial infarction; atherosclerosis;

CC respiratory conditions such as asthma or chronic obstructive pulmonary

CC disease; neoplastic conditions such as cancers or benign tumours; blood

CC cell conditions such as anaemia; gastrointestinal conditions such as

CC gastritis or ulcerative colitis; liver conditions such as biliary

CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;

CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's

CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune

CC system disorders such as acquired immunodeficiency syndrome (AIDS). The

CC nucleic acids may additionally be used to generate animal models of

CC cell death-associated disorders. The present sequence represents a

CC cell death protective polypeptide.

XX

XX Sequence 363 AA;

XX

Query Match 100.0%; Score 37; DB 22; Length 363;

Best Local Similarity 100.0%; Pred. No. 2.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

DB 107 LYQAVATI 114

RESULT 5

AAAB45216

ID AAB45216 standard; Protein: 403 AA.

XX AAB45216;

AC

XX 12-FEB-2001 (first entry)

DT

XX Gene 41 human secreted protein homologous amino acid sequence #157.

DE

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;

KW ophthalmological; autoimmune disease; hyperproliferative disorder;

KW cardiovascular disorder; cerebrovascular disorder; wound healing;

KW nervous system disorder; aging; chemotaxis.

XX Homo sapiens.

OS

XX WO200058467-A1.

PN

XX 05-OCT-2000.

PD

XX 22-MAR-2000; 2000WO-US07505.

XX

XX 26-MAR-1999; 99US-0126502.

PR

XX 17-DEC-1999; 99US-0172410.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-611712/58.

DR

XX Nucleic acid molecules encoding human secreted proteins, used in

PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

PT Parkinson's diseases and cancers -

XX

XX Disclosure; Page 60-61; 440pp; English.

PS

XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human

CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent

CC alternative polypeptides encoded by the genes, and amino acid sequences

CC to which they are homologous. The genes and proteins have activities

CC dependent on the tissues and cells in which they are expressed. Examples

CC of their activities include immunosuppressive; antiarthritic;

CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,  
 CC antagonists and agonists may be useful in treating, preventing and/or  
 CC diagnosing diseases and disorders such as autoimmune diseases  
 CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms  
 CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,  
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous  
 CC system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
 CC The polypeptides can also be used to aid wound healing and epithelial  
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain  
 CC organs before transplantation, for supporting cell culture of primary  
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
 CC also be used as a food additive or preservative to increase or decrease  
 CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences  
 CC used in the isolation and characterisation of the genes and proteins of  
 CC the invention.  
 CC  
 XX  
 SQ Sequence 403 AA;  
 Query Match 100.0%; Score 37; DB 21; Length 403;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYQAVATI 8  
 Db 154 LYQAVATI 161  
 RESULT 6  
 AAB45217  
 ID AAB45217 standard; Protein; 404 AA.  
 XX AAB45217;  
 XX  
 DT 12-FEB-2001 (first entry)  
 DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:158.  
 DE  
 KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;  
 KW ophthalmological; autoimmune disease; hyperproliferative disorder;  
 KW cardiovascular disorder; cerebrovascular disorder; wound healing;  
 KW nervous system disorder; aging; chemotaxis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000058467-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 22-MAR-2000; 2000WO-US07505.  
 XX  
 PR 26-MAR-1999; 99US-0126502.  
 PR 17-DEC-1999; 99US-0172410.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI; 2000-611712/58.  
 DR N-PSDB; AAC80571.  
 XX  
 XX Nucleic acid molecules encoding human secreted proteins, used in  
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 PT Parkinson's diseases and cancers  
 PS  
 PS Disclosure; Page 60-61; 44Opp; English.  
 XX  
 XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human  
 CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent

CC alternative polypeptides encoded by the genes, and amino acid sequences  
 CC to which they are homologous. The genes and proteins have activities  
 CC dependent on the tissues and cells in which they are expressed. Examples  
 CC of their activities include immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,  
 CC antagonists and agonists may be useful in treating, preventing and/or  
 CC diagnosing diseases and disorders such as autoimmune diseases  
 CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms  
 CC of the breast or liver, cardiovascular disorders e.g. cardiac arrest,  
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous  
 CC system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
 CC The polypeptides can also be used to aid wound healing and epithelial  
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain  
 CC organs before transplantation, for supporting cell culture of primary  
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
 CC also be used as a food additive or preservative to increase or decrease  
 CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences  
 CC used in the isolation and characterisation of the genes and proteins of  
 CC the invention.  
 CC  
 XX  
 SQ Sequence 404 AA;  
 Query Match 100.0%; Score 37; DB 21; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYQAVATI 8  
 Db 154 LYQAVATI 161  
 RESULT 7  
 AAG98644  
 ID AAG98644 standard; Protein; 412 AA.  
 XX AAG98644;  
 AC  
 XX 21-SEP-2001 (first entry)  
 DT  
 DE Human cell death protective cDNA clone CNI-00714 ORF1 protein, SEQ:77.  
 DE  
 KW Cell death protective; apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
 KW stroke; cerebral infarction; ischaemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200145638-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 11-DEC-2000; 2000WO-US33547.  
 XX  
 PR 14-DEC-1999; 99US-0461697.  
 XX  
 PR (COGE-) COGENT NEUROSCIENCE INC.  
 PA  
 XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
 PI



XX WPI: 2001-390297/41.  
 DR N-PSDB; AAH84170, AAH84171.  
 XX  
 PT Novel protective sequence polynucleotides and polypeptides, used to  
 PT identify modulators of their expression and activity, which are used in  
 PT to treat central nervous system conditions, diseases and disorders -  
 XX  
 PS Claim 1; Fig 6A; 325pp; English.  
 XX  
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
 CC while the remaining nucleic acid sequences within the range given above  
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
 CC AG98610-AG98829 represent the polypeptides encoded by the cell death  
 CC protective ORFs. The cell death protective cDNA clones are able to  
 CC prevent, delay or reverse progression through the apoptotic or necrotic  
 CC pathways when injected into a cell predisposed to or undergoing cell  
 CC death. The cell death protective nucleic acids and polypeptides can be  
 CC used in the diagnosis and treatment of disorders associated with cell  
 CC death, and to screen for compounds which modulate their activity or  
 CC expression. Such modulators, preferably a small organic molecule, an  
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
 CC cell death-related diseases. Such diseases include those associated with  
 CC the central nervous system including psychiatric or neurological  
 CC disorders, especially ischaemia-related conditions such as strokes, and  
 CC also includes neurodegenerative disorders such as Alzheimer's disease,  
 CC Huntington's disease, or Parkinson's disease. The modulators may also be  
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
 CC vascular diseases such as ischaemic encephalopathy or cerebral  
 CC infarction; eye conditions such as diabetic retinopathy or macular  
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
 CC respiratory conditions such as asthma or chronic obstructive pulmonary  
 CC disease; neoplastic conditions such as cancers or benign tumours; blood  
 CC cell conditions such as anaemia; gastrointestinal conditions such as  
 CC gastritis or ulcerative colitis; liver conditions such as biliary  
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
 CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.  
 XX  
 SQ Sequence 412 AA;

Query Match 100.0%; Score 37; DB 22; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
 |||||  
 Db 156 LYQAVATI 163

RESULT 8  
 AAB03880  
 ID AAB03880 standard; protein; 414 AA.  
 XX  
 AC AAB03880;  
 XX  
 DT 26-OCT-2000 (first entry)  
 XX  
 DE Human ART-1 protein sequence.  
 XX  
 DE Human; tumour antigen protein; ART-1; HLA antigen; cytostatic;  
 XX bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200032770-A1.

XX 08-JUN-2000.  
 PD  
 XX  
 PF 30-NOV-1999; 99WO-JP06682.  
 XX  
 PR 01-DEC-1998; 98JP-0341253.  
 XX  
 PA (SUMU ) SUMITOMO PHARM CO LTD.  
 PA (ITOH/) ITOH K.  
 XX  
 PI Itoh K, Gomi S;  
 XX  
 DR WPI: 2000-412318/35.  
 DR N-PSDB; AAA62864.  
 XX  
 PT Novel tumor antigen protein ART-1, tumor antigen peptide originating  
 PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as  
 PT remedies, preventives and diagnostics for tumors -  
 XX  
 PS Claim 1; Page 44-46; 59pp; Japanese.  
 XX  
 CC The invention relates to a novel human tumour antigen protein, ART-1.  
 CC Included in the invention are polynucleotide sequences encoding the ART-1  
 CC protein, and mutated ART-1 proteins which when broken down  
 CC intracellularly produce a tumour antigen peptide that can recognise HLA  
 CC antigen and bound cytotoxic T cells. Antibodies which specifically  
 CC recognise ART-1 and its derivative peptides, are also included in the  
 CC invention. ART-1 exhibits cytostatic activity. The tumour antigen  
 CC protein, tumour antigen peptide originating from it, their derivatives,  
 CC and DNAs are applicable in vivo or in vitro as remedies, preventives and  
 CC diagnostics for tumours.  
 CC The present sequence represents the human ART-1 protein sequence.  
 XX  
 SQ Sequence 414 AA;

Query Match 100.0%; Score 37; DB 21; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
 |||||  
 Db 158 LYQAVATI 165

RESULT 9  
 AAM25845  
 ID AAM25845 standard; protein; 419 AA.  
 XX  
 AC AAM25845;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:1360.  
 XX  
 DE Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiaggregant; haemostatic; vulnery; antilulcer; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153455-A2.

```

XX PD 26-JUL-2001.
XX PF
XX PR
XX PR 22-DEC-2000; 2000WO-US35017.
XX PR
XX PR 23-DEC-1999; 99US-0471275.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PA (HYSE-) HYSEQ INC.
XX PI
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-457603/49.
XX DR N-PSDB; AAH99786.
XX PT Isolated human polynucleotides encoding polypeptides, useful for the
XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX PS Claim 20; Page 281; 1217pp; English.
XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX CC AAM25963. The proteins can have activities based on the tissues and
XX CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
XX CC central nervous system; viricide; anti-HIV; fungicide; antimutagen;
XX CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
XX CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
XX CC antidiabetic; cytotatic; neuroprotective; antidepressant; nootropic;
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine
XX CC production. The proteins and polynucleotides are useful for screening for
XX CC agonists or antagonists of a protein and for the treatment and diagnosis
XX CC of disorders associated with the activity of a protein e.g. inflammation,
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX CC neurological disorders.
XX SQ Sequence 419 AA;
Query Match 100.0%; Score 37; DB 22; Length 419;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LYQAVATI 8
Db 163 LYQAVATI 170
RESULT 10
ABP28637
ID ABP28637 standard; Protein; 100 AA.
XX AC
XX AC ABP28637;
XX DT
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 6450.
XX KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PF
XX PF 02-MAY-2002.

```

```

XX 29-OCT-2001; 2001WO-GB04789.
XX PF
XX PR
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI
XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN69268.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein -
XX PS Claim 1; Page 3807; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX SQ Sequence 100 AA;
Query Match 86.5%; Score 32; DB 23; Length 100;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LYQAVATI 8
Db 76 LYQEVATI 83
RESULT 11
AAV90939
ID AAV90939 standard; Protein; 273 AA.
XX AC
XX AC AAV90939;
XX DT
XX DT 30-AUG-2000 (first entry)
XX DE Cenarchaeum symbiosum open reading frame protein sequence SEQ ID NO:56.
XX KW Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
XX KW characterisation; archae; therapeutic; industrial; laboratory.
XX OS Cenarchaeum symbiosum.
XX PN WO200018909-A2.
XX PD
XX PD 06-APR-2000.
XX PF 29-SEP-1999; 99WO-US22752.
XX PF 29-SEP-1998; 98US-0102294.
XX PR

```

XX	(DIVE-) DIVERSA CORP.
PA	Swanson RV, Feldman RA, Schleper C;
XX	WPI; 2000-293148/25.
XX	N-PSDB; AAA55214.
DR	New nucleic acids and proteins isolated from the non-thermophilic
XX	crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT	physiology of these archae and in therapeutic, industrial or laboratory
PT	techniques -
XX	Claim 26; Page 171; 210pp; English.
PS	AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
XX	and proteins isolated from the non-thermophilic crenarchaeote
CC	Cenarchaeum symbiosum. The nucleic acids and proteins identified in
CC	the present invention are useful in characterising the physiology of
CC	these archae and can be used in therapeutic, industrial or laboratory
CC	techniques. AAA55227 to AAA55260 represent promoter sequences from
CC	Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
CC	probes used in examples from the present invention.
CC	Sequence 273 AA;
XX	Query Match 86.5%; Score 32; DB 21; Length 273;
CC	Best Local Similarity 87.5%; Pred. No. 28;
CC	Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 LYQAVATI 8
Db	178 LYQAVPTI 185
DE	
DE	-----
RESULT 12	
AAY90923	ID AAY90923 standard; Protein; 269 AA.
XX	AC AAY90923;
XX	DT 30-AUG-2000 (first entry)
XX	Cenarchaeum symbiosum open reading frame protein sequence SEQ ID NO:24.
XX	Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW	characterisation; archae; therapeutic; industrial; laboratory.
XX	Cenarchaeum symbiosum.
OS	WO200018909-A2.
XX	PN 06-APR-2000.
PD	XX
XX	29-SEP-1999; 99WO-US22752.
Pf	XX
XX	29-SEP-1998; 98US-0102294.
PR	XX
XX	(DIVE-) DIVERSA CORP.
PA	Swanson RV, Feldman RA, Schleper C;
PI	XX
XX	WPI; 2000-293148/25.
DR	N-PSDB; AAA55198.
XX	New nucleic acids and proteins isolated from the non-thermophilic
PT	crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT	physiology of these archae and in therapeutic, industrial or laboratory
PT	techniques -
XX	Claim 26; Page 133; 210pp; English.
PS	AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
CC	



DE 2610524B01RIK protein.  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez L., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL: AK012134; BAB28053.1; -;  
DR MGD: MGI:1919445; 2610524B01RIK.  
SQ SEQUENCE 412 AA; 45970 MW; B7BF1525BB44F6D5 CRC64;  
  
Query Match 100.0%; Score 37; DB 11; Length 412;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LYQAVATI 8  
Db 156 LYQAVATI 163  
  
RESULT 3  
O94864 PRELIMINARY; PRT; 414 AA.  
AC O94864;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE KIAA0764 protein (Adenocarcinoma antigen ART1).  
GN KIAA0764.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99087487; PubMed=9872452;  
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,  
RA Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:277-286(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20441578; PubMed=10987294;  
RA Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichijo S.;  
RT "A new tumor-rejection antigen recognized by cytotoxic T lymphocytes  
RT infiltrating into a lung adenocarcinoma.";  
RL Cancer Res. 60:4830-4837(2000).

DR EMBL: AB018307; BAA34484.1; -;  
DR EMBL: AF197954; AAG28523.1; -;  
SQ SEQUENCE 414 AA; 46192 MW; 59724A96353D44D5 CRC64;  
  
Query Match 100.0%; Score 37; DB 4; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LYQAVATI 8  
Db 158 LYQAVATI 165  
  
RESULT 4  
Q9Z1A9 PRELIMINARY; PRT; 891 AA.  
AC Q9Z1A9;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE BUB2-like protein 1 (Vascular RAB-GAP/TBC-containing).  
GN VRP OR HBLP1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEMATOPOIETIC;  
RX MEDLINE=96017645; PubMed=7588067;  
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,  
RA Lee F., McClanahan T.;  
RT "A new approach to the study of haematopoietic development in the yolk  
RT sac and embryoid bodies.";  
RL Development 121:3335-3346(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEMATOPOIETIC;  
RA Guimaraes M.J., Bazan J.F.;  
RT "The CHES protein domain: a novel structural unit that is common  
RT among networks involved in cell signaling, chromatin metabolism and  
RT cell division.";  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U88873; AAD00658.1; -;  
DR EMBL: BC005421; AAH05421.1; -;  
DR MGD: MGI:1927225; Vrp.  
DR InterPro: IPR004182; GRAM\_dom.  
DR InterPro: IPR000195; RabGAP\_TBC.  
DR Pfam: PF02893; GRAM; 1.  
DR Pfam: PF00566; TBC; 1.  
DR SMART: SM00164; TBC; 1.  
SQ SEQUENCE 891 AA; 101976 MW; 394A12B904D06899 CRC64;  
  
Query Match 94.6%; Score 35; DB 11; Length 891;  
Best Local Similarity 75.0%; Pred. No. 18;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 LYQAVATI 8  
Db 769 LYQAVATI 776  
  
RESULT 5  
O95759 PRELIMINARY; PRT; 897 AA.  
AC O95759;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:46:58 ; Search time 1.74408 Seconds  
(without alignments)  
945.129 Million cell updates/sec

Title: US-09-857-308-5  
Perfect score: 37  
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp-archaea.\*  
2: sp-bacteria.\*  
3: sp-fungi.\*  
4: sp-human.\*  
5: sp-invertebrate.\*  
6: sp-mammal.\*  
7: sp-mhc.\*  
8: sp-organelle.\*  
9: sp-phage.\*  
10: sp-plant.\*  
11: sp-rodent.\*  
12: sp-virus.\*  
13: sp-vertebrate.\*  
14: sp-unclassified.\*  
15: sp-rviro.\*  
16: sp-bacterioph.\*  
17: sp-archaeop.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37	100.0	412	4	Q9H2T6
2	37	100.0	412	11	Q9CZV5
3	37	100.0	414	4	Q94864
4	35	94.6	891	11	Q921A9
5	35	94.6	897	4	Q95759
6	32	86.5	100	16	Q9A065
7	32	86.5	273	1	O74043
8	32	86.5	485	17	Q8T1S0
9	32	86.5	865	3	Q9U081
10	31	83.8	269	1	O74066
11	30	81.1	129	15	O36923
12	30	81.1	129	15	Q9QFQ8
13	30	81.1	129	15	Q9QFQ7
14	30	81.1	129	15	Q9QFQ1
15	30	81.1	130	15	O36832
16	30	81.1	130	15	Q9QFQ2

17	30	81.1	130	15	Q9QFQ0	Q9qfq0 human immun
18	30	81.1	130	15	Q9QFQ1	Q9qf1 human immun
19	30	81.1	130	15	Q9QFQ0	Q9qf0 human immun
20	30	81.1	130	15	Q905Y5	Q905y5 human immun
21	30	81.1	130	15	Q905Y4	Q905y4 human immun
22	30	81.1	130	15	Q905Y1	Q905y1 human immun
23	30	81.1	131	15	Q9QFT7	Q9qft7 human immun
24	30	81.1	131	15	Q9QFT6	Q9qft6 human immun
25	30	81.1	131	15	Q9QFQ2	Q9qf2 human immun
26	30	81.1	132	15	Q9QFQ3	Q9qf3 human immun
27	30	81.1	132	15	Q9QFQ1	Q9qf1 human immun
28	30	81.1	132	15	Q9QFQ0	Q9qf0 human immun
29	30	81.1	132	15	Q9QFT9	Q9qft9 human immun
30	30	81.1	132	15	Q9QFT5	Q9qft5 human immun
31	30	81.1	133	15	Q9QG05	Q9qg05 human immun
32	30	81.1	133	15	Q9QG04	Q9qg04 human immun
33	30	81.1	133	15	Q9QG03	Q9qg03 human immun
34	30	81.1	133	15	Q9QG02	Q9qg02 human immun
35	30	81.1	133	15	Q9QG01	Q9qg01 human immun
36	30	81.1	133	15	Q9QFQ4	Q9qf4 human immun
37	30	81.1	133	15	Q9QFQ2	Q9qf2 human immun
38	30	81.1	133	15	Q9QFT8	Q9qft8 human immun
39	30	81.1	133	15	Q9QFS4	Q9qfs4 human immun
40	30	81.1	133	15	Q9QFS3	Q9qfs3 human immun
41	30	81.1	133	15	Q9QFS2	Q9qfs2 human immun
42	30	81.1	133	15	Q9QFS1	Q9qfs1 human immun
43	30	81.1	133	15	Q9QFS0	Q9qfs0 human immun
44	30	81.1	133	15	Q9QFR9	Q9qfr9 human immun
45	30	81.1	133	15	Q9QFR8	Q9qfr8 human immun

ALIGNMENTS

RESULT 1

Q9H2T6 PRELIMINARY; PRT; 412 AA.  
AC Q9H2T6: 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Antigen ART1/P17.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ADENOCARCINOMA;  
RX MEDLINE=20441578; PubMed=10987294;  
RA Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichijo S.;  
RT "A new tumor-rejection antigen recognized by cytotoxic T lymphocytes  
infiltrating into a lung adenocarcinoma.";  
RL Cancer Res. 60:4830-4837(2000).  
DR EMBL: AF224759; AAG47636.1; -;  
SQ SEQUENCE 412 AA; 45950 MW; 99A3830B2E4D3E22 CRC64;

Query Match 100.0%; Score 37; DB 4; Length 412;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

Db 156 LYQAVATI 163

RESULT 2

Q9CZV5 PRELIMINARY; PRT; 412 AA.  
ID Q9CZV5  
AC Q9CZV5;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

```

DE Vascular Rab-GAP/TBC-containing protein.
GN VRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RX MEDLINE=99303817; PubMed=10373574;
RA Yonekura H., Migita H., Sakurai S., Wang H., Harada S., Abedin M.J.,
RT Yamagishi S., Yamamoto H.;
RT "Antisense display-a method for functional gene screening: evaluation
RT in a cell-free system and isolation of angiogenesis-related genes.";
RL Nucleic Acids Res. 27:2591-2600(1999).
DR EMBL: AB024057; BAA75489.1; -.
DR InterPro: IPR004182; GRAM_dom.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF02893; GRAM; 1.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
SQ SEQUENCE 897 AA; 102676 MW; D498C280A657BD1C CRC64;

Query Match 94.6%; Score 35; DB 4; Length 897;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 770 LYQAVATV 777

RESULT 6
ID Q9A065 PRELIMINARY; PRT; 100 AA.
AC Q9A065;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY0914.
GN SPY0914.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006540; AAK3830.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 100 AA; 12041 MW; F06A809D672C5D6C CRC64;

Query Match 86.5%; Score 32; DB 16; Length 100;
Best Local Similarity 87.5%; Pred. No. 8.5;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 76 LYQEVATI 83

RESULT 7
ID O74043 PRELIMINARY; PRT; 273 AA.
AC O74043;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 31.1 kDa protein.
OS Cenarchaeum symbiosum.
OC Archaea; Crenarchaeota; Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RX MEDLINE=98422450; PubMed=9748430;
RA Schlepier C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RT Swanson R.V.;
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL: AF083071; AAC62686.1; -.
DR InterPro: IPR003115; ParBc.
DR InterPro: IPR004437; ParB_part.
DR Pfam: PF02195; ParBc; 1.
DR SMART: SM00470; ParBc; 1.
DR TIGRFAMs: TIGR00180; parB_part; 1.
KW Hypothetical protein.
SQ SEQUENCE 273 AA; 31103 MW; 683C683F18DF61BE CRC64;

Query Match 86.5%; Score 32; DB 1; Length 273;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 178 LYQAVPTI 185

RESULT 8
ID Q8TLS0 PRELIMINARY; PRT; 485 AA.
AC Q8TLS0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Proton/sodium-glutamate symporter.
GN MA2961.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010998; AAM06334.1; -.
KW Complete proteome.
SQ SEQUENCE 485 AA; 51158 MW; 4A3C504F95A1B322 CRC64;

Query Match 86.5%; Score 32; DB 17; Length 485;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8
Db 111 LYQAVATI 111

```

Db 348 LYQGVATV 355

RESULT 9

Q9U081 PRELIMINARY; PRT; 865 AA.

AC Q9U081; 174 LYQAVPTV 181

DT 01-MAY-2000 (TREMELrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Putative gamma-adaptin.

SPCPIEL1.06.

GN Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972H-;

RA Wedler H., Duetherhoeft A., McDougall R.C., Rajandream M.A.,

RA Barrell B.G.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL01183; CAB54865.1; -

DR InterPro; IPR002553; Adaptin\_N.

DR InterPro; IPR001121; Gamma\_adaptin\_C.

DR Pfam; PF01602; Adaptin\_N; 1.

DR Pfam; PF02883; Alpha\_adaptin\_C2; 1.

DR ProDom; PD021457; Gamma\_adaptin\_C; 1.

SQ SEQUENCE 865 AA; 96041 MW; 05BA43A15B61E84E CRC64;

Query Match 86.5%; Score 32; DB 3; Length 865;

Best Local Similarity 87.5%; Pred. No. 92;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

Db 317 LYQAVRTI 324

RESULT 10

O74066 PRELIMINARY; PRT; 269 AA.

AC O74066;

DT 01-NOV-1998 (TREMELrel. 08, Created)

DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Hypothetical 30.7 kDa protein.

OS Cenarchaeum symbiosum.

OC Archaea; Crenarchaeota; Cenarchaeum.

OX NCBI\_TaxID=46770;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=B;

RX MEDLINE=98422450; PubMed=9748430;

RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,

RA Swanson R.V.;

RT "Genomic analysis reveals chromosomal variation in natural populations

RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";

RL J. Bacteriol. 180:5003-5009(1998).

DR EMBL; AF083072; AAC62709.1; -

DR InterPro; IPR003115; ParBc.

DR InterPro; IPR004437; ParB\_part.

DR Pfam; PF02195; ParBc; 1.

DR SMART; SM00470; ParBc; 1.

DR TIGRfams; TIGR00180; parB\_part; 1.

KW Hypothetical protein.

SQ SEQUENCE 269 AA; 30689 MW; AA45216CAD16DD5B CRC64;

Query Match 83.8%; Score 31; DB 1; Length 269;

Best Local Similarity 75.0%; Pred. No. 44;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

Db 174 LYQAVPTV 181

RESULT 11

O36923 PRELIMINARY; PRT; 129 AA.

ID O36923;

AC O36923;

DT 01-JAN-1998 (TREMELrel. 05, Created)

DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)

DE GAG polyprotein (Fragment).

GAG.

GN Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=97445059; PubMed=9300048;

RA Leigh Brown A.J., Lobidel D., Wade C.M., Rebus S., Phillips N.,

RA Brettle R.P., France A.J., Leen C.S., McMenamin J., McMillan A.,

RA Maw R.D., Mulcahy F., Robertson J.R., Sankar K.N., Scott G., Wyld R.,

RA Peutherer J.F.;

RT "The molecular epidemiology of human immunodeficiency virus type 1 in

RT six cities in Britain and Ireland.";

RL Virology 235:166-177(1997)

DR EMBL; AF014320; AAC58401.1; -

DR InterPro; IPR000071; Retrovir\_p17.

DR Pfam; PF00540; Gag\_p17; 1.

KW AIDS; Core protein; Polyprotein.

FT NON\_TER 1

FT NON\_TER 129

SQ SEQUENCE 129 AA; 14468 MW; 22D4B58167285DF9 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 129;

Best Local Similarity 75.0%; Pred. No. 34;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8

Db 55 LYNAVATL 62

RESULT 12

Q9QF08 PRELIMINARY; PRT; 129 AA.

ID Q9QF08;

AC Q9QF08;

DT 01-MAY-2000 (TREMELrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)

DE GAG polyprotein (Fragment).

GAG.

GN Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LN23;

RX MEDLINE=99412391; PubMed=10482626;

RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettell R.P.,

RA Bell J.E., Simmonds P.;

RT "Mosaic structure of the human immunodeficiency virus type 1 genome

RT infecting lymphoid cells and the brain: evidence for frequent in vivo

RT recombination events in the evolution of regional populations.";

RL J. Virol. 73:8720-8731(1999).

DR EMBL; AF174940; AAF00286.1; -

DR InterPro; IPR000071; Retrovir\_p17.

DR Pfam; PF00540; Gag\_p17; 1.

DR PRINTS; PR00234; HIVMATRIX.

KW AIDS; Core protein; Polyprotein.

FT NON\_TER 1

FT NON\_TER 129



```
SQ SEQUENCE 129 AA; 14574 MW; C24B7A484562E0 CRC64;
Query Match 81.1%; Score 30; DB 15; Length 129;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 51 LYNAVATL 58

RESULT 13
Q9QFQ7 PRELIMINARY; PRT; 129 AA.
AC Q9QFQ7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GAG polyprotein (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LN24;
RX MEDLINE=99412391; PubMed=10482626;
RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettle R.P.,
RA Bell J.E., Simmonds P.;
RT "Mosaic structure of the human immunodeficiency virus type 1 genome
RT infecting lymphoid cells and the brain: evidence for frequent in vivo
RT recombination events in the evolution of regional populations.";
RL J. Virol. 73:8720-8731(1999).
DR EMBL: AF174941; AAF00287.1; -
DR InterPro: IPR000071; Retrovir_p17.
DR Pfam: PF00540; Gag_p17; 1.
DR PRINTS; PR00234; HIV1MATRIX.
DR AIDS; Core protein; Polyprotein.
KW NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14448 MW; C2234C7181D896B3 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 129;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 51 LYNAVATL 58

RESULT 14
Q9QFQ1 PRELIMINARY; PRT; 129 AA.
AC Q9QFQ1;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GAG polyprotein (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LN5;
RX MEDLINE=99412391; PubMed=10482626;
RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettle R.P.,
RA Bell J.E., Simmonds P.;
RT "Mosaic structure of the human immunodeficiency virus type 1 genome
RT infecting lymphoid cells and the brain: evidence for frequent in vivo
RT recombination events in the evolution of regional populations.";
RL J. Virol. 73:8720-8731(1999).
```

```
DR EMBL: AF174947; AAF00293.1; -
DR InterPro: IPR000071; Retrovir_p17.
DR Pfam: PF00540; Gag_p17; 1.
DR PRINTS; PR00234; HIV1MATRIX.
KW AIDS; Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14474 MW; 007EF7119CFFEF41 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 129;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 51 LYNAVATL 58

RESULT 15
Q36832 PRELIMINARY; PRT; 130 AA.
AC Q36832;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GAG polyprotein (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97445059; PubMed=93000048;
RA Leigh Brown A.J., Lobidel D., Wade C.M., Rebus S., Phillips N.,
RA Brettle R.P., France A.J., Leen C.S., McMenamin J., McMillan A.,
RA Maw R.D., Mulcahy F., Robertson J.R., Sankar K.N., Scott G., Wyld R.,
RA Peutherer J.F.;
RT "The molecular epidemiology of human immunodeficiency virus type 1 in
RT six cities in Britain and Ireland.";
EL Virology 235:166-177(1997).
DR EMBL: AF014229; AAC58310.1; -
DR InterPro: IPR000071; Retrovir_p17.
DR Pfam: PF00540; Gag_p17; 1.
DR PRINTS; PR00234; HIV1MATRIX.
KW AIDS; Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 130
SQ SEQUENCE 130 AA; 14450 MW; E7AA32546F1B2656 CRC64;

Query Match 81.1%; Score 30; DB 15; Length 130;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 55 LYNAVATL 62

Search completed: July 18, 2003, 09:59:38
Job time : 2.74408 secs
```

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:46:58 ; Search time 90.2559 Seconds  
(without alignments)  
945.129 Million cell updates/sec

Title: US-09-857-308-1  
Perfect score: 2175  
Sequence: 1 MNLQRYWGEIPISSTNRS.....SLMGSSPVFNQCKRMRKI 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2175	100.0	414	4	Q94864
2	2155	99.1	412	4	Q9H2T6
3	2079	95.6	412	11	Q9CZV5
4	142.5	6.6	979	3	P87152
5	132	6.1	359	5	Q9VX12
6	131	6.0	1049	5	Q9V149
7	126	5.8	1621	16	Q9KTA5
8	121.5	5.6	505	3	Q8X0V8
9	121.5	5.6	2957	5	Q61845
10	121.5	5.6	2977	5	Q9VAP9
11	120.5	5.5	893	11	Q99P15
12	118.5	5.4	582	11	Q9ER99
13	117	5.4	335	4	Q9Y4V2
14	117	5.4	486	4	Q96HV9
15	116.5	5.4	791	11	Q9Z1P7
16	114.5	5.3	17352	5	Q95YM2

17	114	5.2	486	11	Q9QV18
18	113	5.2	1152	4	Q92603
19	113	5.2	1204	4	Q00211
20	113	5.2	1235	4	Q9H2G2
21	113	5.2	1852	5	Q9VH42
22	112	5.1	486	4	Q9UNF0
23	112	5.1	1278	4	Q9UPP5
24	111.5	5.1	660	5	Q9VT63
25	111.5	5.1	1248	3	Q8X0V7
26	109.5	5.0	6815	5	Q917U4
27	109.5	5.0	16215	5	Q9NFS3
28	109	5.0	486	11	Q9WVE8
29	109	5.0	678	10	Q94I68
30	108.5	5.0	632	4	Q94937
31	108.5	5.0	1206	11	Q08815
32	108.5	5.0	1700	5	Q9VVA9
33	108	5.0	1266	5	Q9GRH4
34	107.5	4.9	568	12	Q9DUE2
35	107.5	4.9	3263	5	Q917U3
36	107	4.9	884	5	Q61543
37	107	4.9	884	5	Q9VEV9
38	107	4.9	2484	5	Q9U347
39	107	4.9	2607	5	Q23187
40	106	4.9	599	10	Q93ZU5
41	106	4.9	599	10	Q9LRR7
42	106	4.9	611	5	Q16938
43	106	4.9	1180	5	Q18673
44	106	4.9	4667	5	Q9TVI9
45	106	4.9	5170	5	Q17490

#### ALIGNMENTS

#### RESULT 1

ID	Q94864	PRELIMINARY;	PRT;	414 AA.
AC	Q94864;			
DT	01-MAY-1999 (TrEMBLrel. 10, Created)			
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)			
DE	KIAA0764 protein (Adenocarcinoma antigen ART1).			
GN	KIAA0764.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RX	MEDLINE=99087487; PubMed=9872452;			
RA	Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,			
RA	Kotani H., Nomura N., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. XI.			
RT	The complete sequences of 100 new cDNA clones from brain which code			
RT	for large proteins in vitro."			
RL	DNA Res. 5:277-286(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20441578; PubMed=10987294;			
RA	Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichiho S.;			
RT	"A new tumor-rejection antigen recognized by cytotoxic T lymphocytes			
RT	infiltrating into a lung adenocarcinoma."			
RL	Cancer Res. 60:4830-4837(2000).			
DR	EMBL; AB018307; BAA34484.1; -.			
DR	EMBL; AF197954; AAG28523.1; -.			
SQ	SEQUENCE 414 AA; 46192 MW; 59724A96353D44D5 CRC64;			
Query Match 100.0%; Score 2175; DB 4; Length 414;				
Best Local Similarity 100.0%; Pred. No. 1.8e-166;				
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				

QY 1 MNLQRYWGEIPISSTNRSFDLLPREFLVEVHDPPLHQPSSANKPKPTMLDIPSEPC 60

Db 1 MNLQRYWGEIPISSQTNRSFDLLPREFLVEVDPLHQPSANKPKPTMLDIPSEPC 60  
 Qy 61 SLTHTIQLIQRNRLNLATAQAQNOOQTEGVKTESEPLPSCPGSPPLDPLDLC 120  
 Db 61 SLTHTIQLIQRNRLNLATAQAQNOOQTEGVKTESEPLPSCPGSPPLDPLDLC 120  
 Qy 121 KNPAPFOIRHSDPESDFYRGKGEFVTELSWHSCRQLLYQAVATILAHAGFDCANESYLE 180  
 Db 121 KNPAPFOIRHSDPESDFYRGKGEFVTELSWHSCRQLLYQAVATILAHAGFDCANESYLE 180  
 Qy 181 TLTDVAHEYCYLKFLLRFVAVDREARLGQTFFPDVMEQVFEVGVGSLQKFWQHRKI 240  
 Db 181 TLTDVAHEYCYLKFLLRFVAVDREARLGQTFFPDVMEQVFEVGVGSLQKFWQHRKI 240  
 Qy 241 DYHSYMLQISKOLSEERYIVNPEKATEDAKPVKIKEEVPVSDITFFPVEEELADLASGQ 300  
 Db 241 DYHSYMLQISKOLSEERYIVNPEKATEDAKPVKIKEEVPVSDITFFPVEEELADLASGQ 300  
 Qy 301 SLPNGVLGAQSERFPPSNLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNYSGHVL 360  
 Db 301 SLPNGVLGAQSERFPPSNLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNYSGHVL 360  
 Qy 361 GSDVFEPMGMSGSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKKMRKI 414  
 Db 361 GSDVFEPMGMSGSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKKMRKI 414

RESULT 2  
 09H2T6

Db 243 YMLQISKOLSEERYIVNPEKATEDAKPVKIKEEVPVSDITFFPVEEELADLASGQSLPM 302  
 Qy 305 GVLGAQSERFPPSNLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNYSGHVGLSDV 364  
 Db 303 GVLGAQSERFPPSNLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNYSGHVGLSDV 362  
 Qy 365 FEPMGMSGSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKKMRKI 414  
 Db 363 FEPMGMSGSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKKMRKI 412

RESULT 3  
 09CZV5  
 ID Q9CZV5 PRELIMINARY; PRT; 412 AA.  
 AC Q9CZV5;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE 2610524B01RIK protein.  
 GN 2610524B01RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK012134; BAB28053.1; -;  
 DR MGD; MGI:1919445; 2610524B01RIK.  
 SQ SEQUENCE 412 AA; 45970 MW; B7BF1525BB44F6D5 CRC64;

Query Match 95.68; Score 2079; DB 11; Length 412;  
 Best Local Similarity 95.9%; Pred. No. 9.2e-159;  
 Matches 393; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 5 RYWGEPISQTNRSFDLLPREFLVEVDPLHQPSANKPKPTMLDIPSEPCSLTI 64  
 Db 3 RYWGEPISQTNRSFDLLPREFLVEVDPLHQPSANKPKPTMLDIPSEPCSLTI 62  
 Qy 65 HTTOLIOHNRRLNLATAQAQNOOQTEGVKTESEPLPSCPGSPPLDPLDCKNPN 124  
 Db 63 HTTOLIOHNRRLNLATAQAQNOOQTEGVKTESEPLPSCPGSPPLDPLDCKNPN 122  
 Qy 125 APFOIRHSDPESDFYRGKGEFVTELSWHSCRQLLYQAVATILAHAGFDCANESVLETLTD 184  
 Db 123 APFOIRHSDPESDFYRGKGEFVTELSWHSCRQLLYQAVATILAHAGFDCANESVLETLTD 182  
 Qy 185 VAHEYCYLKFLLRFVAVDREARLGQTFFPDVMEQVFEVGVGSLQKFWQHRKIDYHS 244  
 Db 183 VAHEYCYLKFLLRFVAVDREARLGQTFFPDVMEQVFEVGVGSLQKFWQHRKIDYHT 242  
 Qy 245 YMLQISKOLSEERYIVNPEKATEDAKPVKIKEEVPVSDITFFPVEEELADLASGQSLPM 304

Db 63 HTTOLIOHNRRLNLATAQAQNOOQTEGVKTESEPLPSCPGSPPLDPLDCKNPN 122  
 Qy 125 APFOIRHSDPESDFYRGKGEFVTELSWHSCRQLLYQAVATILAHAGFDCANESVLETLTD 184  
 Db 123 APFOIRHSDPESDFYRGKGEFVTELSWHSCRQLLYQAVATILAHAGFDCANESVLETLTD 182  
 Qy 185 VAHEYCYLKFLLRFVAVDREARLGQTFFPDVMEQVFEVGVGSLQKFWQHRKIDYHS 244  
 Db 183 VAHEYCYLKFLLRFVAVDREARLGQTFFPDVMEQVFEVGVGSLQKFWQHRKIDYHS 242  
 Qy 245 YMLQISKOLSEERYIVNPEKATEDAKPVKIKEEVPVSDITFFPVEEELADLASGQSLPM 304

```
Db 243 YMLQISKQLEEEYRINPKATDTPVKIKKEPVSDITFPVSEELADLASQDQLPI 302
QY 305 GVLCAQSERFNSLEVEASQASSAEVNASPLNLHVKMEPOSEBGNVSGHVLGSDV 364
Db 303 GVLCAQSERFNSLEVEASQASSAEVNASPLNLHVKMEPOSEBGNVSAHVLGSDV 362
QY 365 FEPMSCMSBAGIQSPDDSDSYSGHSTDSLMSGSPVFNQRCCKRMKI 414
Db 363 FEPMSCMSBAGLQSPDDSDSYSGHSTDSLMSGSPVFNQRCCKRMKI 412

RESULT 4
P87152 PRELIMINARY; PRT; 979 AA.
ID P87152
AC P87152;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcriptional activator C25H2.11C.
GN SPBC25H2.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Darso G., Iye G., Bowman S., Church C., Wood V., Barrell B.G.,
RA Rajadream M.A., Connor R.E.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS A COPY OF THE BROMODOMAIN.
DR EMBL: 295397; CAB08777.1; -.
DR HSP: Q92831; B991.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 1.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN.1; 1.
DR PROSITE: PS50014; BROMODOMAIN.2; 1.
KW Hypothetical protein; Transcription regulation; Nuclear protein;
KW Activator; Bromodomain.
FT DOMAIN 316 386 BROMODOMAIN (BY SIMILARITY).
SQ SEQUENCE 979 AA; 111503 MW; D60E78663B8C68AF CRC64;

Query Match 6.6%; Score 142.5; DB 3; Length 979;
Best Local Similarity 19.4%; Pred. No. 0.011;
Matches 83; Conservative 74; Mismatches 183; Indels 87; Gaps 14;

QY 10 IPISSQT--NRSSFLLPREFLVEHPDPLHOPANKPKPTMLDIPSEPCSLTIHTI 67
Db 596 VPMTAGSVLESEEDLYFRDYSLFEI-----NRNTPGV-----SLMKNI 637
QY 68 QLTQHNRRNLNLIATAQNOQTEGVKTESEPLSCPSPLPDLPLCKNPAPF 127
Db 638 AKMQETKRLCKNIQTQLQLPQPFYEHKSH-VFPANNEPILLD--IPQNYDNSSPK 694
QY 128 QIRHSDPESFYKGEPTELSWHSCROLLYQAVATILAHAGFCANESVLETLTDVAH 187
Db 695 PLAH-----DVLKLCILIFLHAGFESFQMGALDALTEIAA 730
QY 188 EYCLK----FTKLLREAVDREARLGQTFFPDVMEQVFEHVGISVLSLQKFMQHRKDYH 243
Db 731 DYNAKMGAVMDQYLIYKDKSQ-----EIVGQTLGELGVDDVNDLISYVYHVEROS 783
QY 244 SYMLQISKQLEEEYRINPKATDTPVKIKKEPVSDITFPVSE-----L 291
Db 784 VKLLETHQRLQRHFVLLRPAERNDDEAIFNQGESFVTGNFSETGDDFFGLRELGL 843
QY 292 EADLASGDSPLMGVLCGASERPPSNLEVEASQAS--SAEVNASPLNLHVKMEPOSE 350
Db 844 DRELGLDLSVLPHLL---QSLRSLNKSQW---PEATIKGDQEVAPPKPPYPTAESISNE 898
```

```
QY 351 EGNVSGHGVLSGSDVF-----EPPMSGMSEAGIQSPDDSDSYSGHSTDSLMSGSPVF 403
Db 899 IGLIQGLFLKLNLEFGLDELLEDEIRPSKPPRPLPNGKITTR-----KRIASSVFL 954
QY 404 NQRCKKR 410
Db 955 NQSLRKK 961

RESULT 5
Q9VX12 PRELIMINARY; PRT; 359 AA.
ID Q9VX12
AC Q9VX12;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG6506 protein.
GN CG6506.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003506; AAF48771.1; -.
DR FlyBase: FBgn0030874; CG6506.
SQ SEQUENCE 359 AA; 39460 MW; 26440857AFAD6854 CRC64;
```

Query Match 6.1%; Score 132; DB 5; Length 359;  
Best Local Similarity 18.7%; Pred. No. 0.018;  
Matches 71; Conservative 71; Mismatches 174; Indels 64; Gaps 12;



RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
RT cholerae";  
RL Nature 406:477-483(2000).  
DR EMBL: AE004181; AAF94159.1; -;  
KW TIGR: VC0998; -;  
DR Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1621 AA; 178073 MW; 2EF38BAC27A2485D CRC64;

Query Match 5.8%; Score 126; DB 16; Length 1621;  
Best Local Similarity 21.7%; Pred. No. 0.46;  
Matches 93; Conservative 75; Mismatches 169; Indels 92; Gaps 20;

QY 9 EIPSSQTNRSFDLL-----PREFRLVEHDPPLHQPSSANKPKPPTMLDIPSPCSLT 63  
DB 815 DLPEQTATNETADELLADLAQPOSNTVDTSDAL-----APDGLSQSVEEP--LT 864  
QY 64 IHTIQLIQRNRLNL-IATAQAOQOQTEGYKTEESEPLPSCPGSP-----PLP 112  
DB 865 LNDLELPEENDEPQLAEVTPSSAFDEQVETEIEPESEPLAAEASNDESDLTALNELDLP 924  
QY 113 ----DDLPLDCKNNAFQIRHSDPESDFYRGKGPVTELSHSCRLLYQA--TTL 166  
DB 925 EYTEEDVLADVQLEPAEAEV---EPDLELV---NEPVTAEAFTELDLDPYTEEDAL 978  
QY 167 AHAGFDCAVESVLETLTDAHECYCLKFKLLRFADVREA--RLGQTFFPDVMEQVFEHVG 224  
DB 979 ADAQLEPAEVAEVEPELDLASE-----PAEEEAFTLNLKLDLPYTEE----- 1021  
QY 225 IGSVLSLOKFWQHRKIDYHMYLQISKOLSE-----EYERIVNPEKATEDA----- 270  
DB 1022 --DALADQALESATSESEVSELELVSEPAEAEAFTELDLDPYTEEDALADSOLEPAA 1079  
QY 271 -----KPKVKEEVPDITFPVSELEADLASGQSLPMGVL--GAQSERFFSNLEVAS 323  
DB 1080 ESEVEPELSEVPEETEAFTLDELDPYTEEDALADQAQLEPAVESEVPE-ELELASE 1138  
QY 324 P-QASSAEVNASPLNN-----LAHVMEPESEEGNVSGHVGLSGV-FEPPMSGNSE 374  
DB 1139 PAEEEAFTLNLDPYTEEDALADQAQLEP--AAEVESESELELASDLEEEPEFTLNE 1196  
QY 375 AGIQSPSD 383  
DB 1197 LDLPEYTEE 1205

RESULT 8  
Q8X0V8 PRELIMINARY; PRT; 505 AA.  
AC Q8X0V8  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)  
DE Hypothetical 55.6 kDa protein.  
GN 12344.340.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL670009; CAD21385.1; -;  
KW Hypothetical protein.

SQ SEQUENCE 505 AA; 55613 MW; 4E88D0BBA42383E7 CRC64;  
Query Match 5.6%; Score 121.5; DB 3; Length 505;  
Best Local Similarity 17.7%; Pred. No. 0.2;  
Matches 76; Conservative 69; Mismatches 184; Indels 101; Gaps 16;

QY 10 IPISSTQNRSSFDLLPREFRLVEHDPPLHQPSSANKPKPPTMLDIPSPCSLTHTITQL 69  
DB 22 VPLPSYADNSSTSSSGRLRSRRPAPPSPPPSPKPKKATSSRAPEPP-----QETVQP 77  
QY 70 IQHNRRLNLATAQAOQOQTEGYKTESEPL---PSCPG--SPL-PPDLLPLDCKNP 123  
DB 78 V-----SPLPHEDEEQPKATPKQMPILRPPTQTGTSPPPQHPDDIYRMPPLPP 126  
QY 124 NAFQIRHSDPESDFYRGKGPVTELSHSCRLLYQA--VATILAHAGFDCAVESVLETL 182  
DB 127 YLSVQPRICEP-ADNYRG-----WTTPLFESYKMNHPLTYGSE----- 165  
QY 183 TDAHECYCLKFKLLRFADVREAARLGQTPFPDME-----QVFHEVGIGS--- 227  
DB 166 -DIWFEQVKMAEASNCSELTNTRICKFPDGPVTEAFICQMRFLSCWRILHGTISGAYWA 224  
QY 228 ---VLSLOKF-----WQHRKIDYHMYLQISKOL---SEEVERIVN 262  
DB 225 YNRVLGYSRIPVFTGAPDGGSGEQSACQWQPSPDCVYFALEVKRVTPVNSEQQPWYQ 284  
QY 263 PEKATEDAKPVKIKEP-----VSDITFPVSELEADLASGQSLPMGVLGAQSERFPSNL 318  
DB 285 TNRWVEEVSRAKTYDPSVQYKQEGMAWLKEVNLGMILKGDQVCESMYDRKPDWPEWP 344  
QY 319 EYEASPOASSAEVNASPLNNLAHVMEPESEEGNVSGHVGLSGVFEPPMSGMSEAGIP 378  
DB 345 EVPAGGE-----VEPPWALGSLRAEDAFSGSGAGVYVSS-----KNAGHIGQQ 391  
QY 379 QSPDSDSSY 388  
DB 392 HQPQQAQSTF 401

RESULT 9  
O61845 PRELIMINARY; PRT; 2957 AA.  
ID O61845  
AC O61845;  
DT 01-AUG-1998 (TremBLrel. 07, Created)  
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)  
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)  
DE T04D1.4 protein.  
GN T04D1.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkison-Sproat J., Wohldmann P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Davidson S., Wohldmann P.;





```
Db 2230 RQLNGKTFQE-----ILMQESTLPHRT-----SELLQMOQLMIDDLKGEINERTAEKELEQ 2280
Qy 274 KIKEEPVSDITFPVSELEADLASDOSLPMGLGAQSERPPSNLEVEASPOA----- 326
Db 2281 QLKKEEQTSE-----KSTAESDVLSEAK-----VKEEIKQEPIDKSAQDGVVDFDL 2327
Qy 327 --SSAEVNASPLMNLNAHVHAKBPQBS-----EEGNVSGHGVLSGVDFPEPMSCMSEAGTIPQ 379
Db 2328 LKNSADVNK-----AFSKSQPQQTIVTKQESNESGEPSPVSKSEGHDKESKYLKK 2381
Qy 380 SPDDSD 385
Db 2382 DYDNFD 2387

RESULT 11
Q99P15 PRELIMINARY; PRT; 893 AA.
AC Q99P15;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE LPIN2.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20578762; PubMed=11138012;
RA Peterfy M., Phan J., Xu P., Reue K.;
RT "Lipodystrophy in the fld mouse results from mutation of a new gene
RT encoding a nuclear protein, lipin.";
RL Nat. Genet. 27:121-124(2001).
DR EMBL; AF286723; AAG52761.1; -.
DR MGI; MGI:1891341; Lpin2.
DR InterPro; IPR001395; Aldo/ket_red.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; UNKNOWN_1.
SQ SEQUENCE 893 AA; 99617 MW; 3ADC0C240626B574 CRC64;

Query Match 5.5%; Score 120.5; DB 11; Length 893;
Best Local Similarity 18.9%; Pred. No. 0.54;
Matches 76; Conservative 52; Mismatches 136; Indels 139; Gaps 16;

Qy 7 WGEIPISSTNRSFOLLPR-----EFLVEVHDPPLHQPAN----- 45
Db 266 WGGFPSTKVKRYDYHPTATITPSENTHFRVIPSEDSLIREVEKDATVEDTTCIV 325
Qy 46 KKPPTW-----LDIPSEPCSLTIHTLIQHNRRLNLIATAQAQNOQOQTEGKVT 96
Db 326 KPRPRALCKQLSDAASTELPEP-----LEAPQISLLDADPVPS 365
Qy 97 EE-----SEPLPSCPGSPPL-----PDDLPLDCK--NPN-APFQIRHS----- 133
Db 366 PSAEAPSEPKPAKDSPTKKXGVHKSQHQPDDIYLLDLKALEPEVAALYFPKSDTPG 425
Qy 134 ----PESDFRGKGPVTELSWISCRQLLYQAVATILAHAGFCANESVLETLTDVAHEY 189
Db 426 SROMPESDTFSGSQSP-----QSVGSAADSGTECLSDAMD-LPDVTLSL 470
Qy 190 C-----LKFTKLLRFADVAREARGOTFPFVMEQVFEH-----VIGISVLSLOK 233
Db 471 CGGLSENGEISKEKFEWHIITYHEFAPNPGLDINPNLVIIRYNNYAWAAMPILSLQV 530
Qy 234 FWQHRINDXYSYMLQISKQSEYERIVNPEKA-----TEDAKPVKIKEE 278
Db 531 FQK-----SLPKATVESWVKDKMPKSGRWFWFKKESMIKQLPETKEGKSEVP 579
Qy 279 PVSDITFPVSELEADLASDOSLPMGLGAQSERPPSNLEVE 321
Db 580 PANDLPSNABEPTSARPAENDTSSDEG-----SQELEESIKVD 617
```

## RESULT 12

```
Q9ER99 PRELIMINARY; PRT; 582 AA.
ID Q9ER99;
AC Q9ER99;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Regulatory subunit of SGLT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Baumgarten K., Gorboulev V., Koepsell H.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y11917; CAA72676.1; -.
DR InterPro; IPR000449; UBA_domain.
DR SMART; SM00165; UBA; 1.
SQ SEQUENCE 582 AA; 61247 MW; 23CB8C50CA2B390F CRC64;

Query Match 5.4%; Score 118.5; DB 11; Length 582;
Best Local Similarity 20.8%; Pred. No. 0.43;
Matches 83; Conservative 52; Mismatches 162; Indels 103; Gaps 18;

Qy 40 HQPSANKPKPTMLDIPSEPCSLTIHTLIQHNRRLNLIATAQAQNOQO----- 90
Db 212 HLPSEAKGLPAS--GLCSCPCSEALMEVDTAEE-----QSLVAMCSSTGRQDAVIKPSVA 264
Qy 91 --TEGVKTESEPLPSCPGSPPLDPLLPLDCKNPNAPFOI-----RHSDPESDFYRGKE 144
Db 265 HLASDNTMETVETLQSNPCEPVEHSILTRQLQPLPEDNVDMSTMDNDDSSLLSGHQ 324
Qy 145 PVTELSWHSCRQLLYQAVATILAHAGFCANESVLETLTDVAHEYCKLFTKLLRFADVRE 204
Db 325 PSVESAEFFC-----SSVTVALKEL-HELLVTSCK----- 353
Qy 205 ARLGQTFPFDVMEQVHF-----EVGIGSVLSLQKFWQHRIKDYH-----SYMLQISKQLS 254
Db 354 -----PASESEPHVTQSEIGAESQPSVDSLGRVRVQSVHLTSPDOYSGSCHOAT 405
Qy 255 EYER--IVN--PKATATEDAKPVKIKBEPVSDITFPVSELEADLASDOSLPMGLGAQ 310
Db 406 SESKTIIVGTAPCAAVEDEASTSF--EGLGDGLSPDREDVRRSTESARKSCSVAITSAK 463
Qy 311 -SERFPNLEVEASPOASSAE-VNASPLMNLNAHV-KMPEOSESNEGNSVGHGVLSGVDFEE 367
Db 464 LSEQLPCTGLGVEIAPELAAASEGHSQP---SERVHNPGDPDPETSSVC----- 508
Qy 368 PMSCMSAGIQSPDDSDSS-----YGSHTSDLSLMS 399
Db 509 PGAGLPRSLDQPPPTQSLSTPSPVLPPIFFPAADVDRILGA 548

RESULT 13
Q9Y4V2 PRELIMINARY; PRT; 335 AA.
ID Q9Y4V2;
AC Q9Y4V2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE D0437M21.3 (Protein kinase C and casein kinase substrate in neurons 2)
DE (Fragment).
GN PACSIN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Laird G.;
```



Search completed: July 18, 2003, 09:59:37  
Job time : 96.2559 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:40:17 ; Search time 2561 Seconds  
(without alignments)  
10820.179 Million cell updates/sec

Title: US-09-857-308-2

Perfect score: 1711

Sequence: 1 acgcgatccttgcctcaggc.....aaaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1057	61.8	2612	11 AK012134	AK012134 Mus muscu
2	940.4	55.0	1027	14 BM921322	BM921322 AGENCOURT
3	904.6	52.9	939	13 BM545491	BM545491 AGENCOURT
4	824.4	48.2	865	14 BQ229669	BQ229669 AGENCOURT
5	817.4	47.8	1214	13 BM545988	BM545988 AGENCOURT
6	797.4	46.6	923	14 BQ924589	BQ924589 AGENCOURT

7	758.4	44.3	904	13	BI822539	BI822539 603035248
8	750.6	43.9	818	13	BI669780	BI669780 603293356
9	738.2	43.1	972	13	BM468753	BM468753 AGENCOURT
10	731.2	42.7	938	12	EG681414	EG681414 602627777
11	721.4	42.2	771	13	BI770385	BI770385 603055622
12	710.8	41.5	735	14	BQ576180	BQ576180 UI-H-E21-
13	706	41.3	821	12	BG489862	BG489862 602518753
14	704.4	41.2	718	14	BM728984	BM728984 UI-E-B01-
15	702	41.0	899	14	BQ887284	BQ887284 AGENCOURT
16	700.6	40.9	788	13	BI763619	BI763619 603050193
17	691.8	40.4	764	13	BI855674	BI855674 603383202
18	686.4	40.1	798	13	BI193419	BI193419 602947051
19	684.2	40.0	725	12	BG149842	BG149842 nae01c08.
20	677.8	39.6	681	13	BI160260	BI160260 602863964
21	665.8	38.9	758	12	BG706749	BG706749 602671875
22	664.4	38.8	712	14	BM989602	BM989602 UI-H-DH0-
23	658.6	38.5	796	9	AL559070	AL559070 AL559070
24	656	38.3	783	13	BI602617	BI602617 603248445
25	654.4	38.2	1059	12	BG323193	BG323193 602421403
26	634.6	37.1	1105	14	BM802810	BM802810 AGENCOURT
27	627.8	36.7	687	13	BM050170	BM050170 603632231
28	626.4	36.6	1059	14	BM802006	BM802006 AGENCOURT
29	615	35.9	616	9	AI130886	AI130886 qb81e05.x
30	606	35.4	839	12	BG163932	BG163932 602343111
31	601.8	35.2	688	13	BG927573	BG927573 HNC43-1-C
32	596	34.8	608	12	BG776075	BG776075 602663014
33	594	34.7	594	9	AA694172	AA694172 z142a01.s
34	594	34.7	598	13	BM310780	BM310780 i48d11.y
35	592.8	34.6	619	13	BI907412	BI907412 603063781
36	592.8	34.6	679	13	BG927465	BG927465 HNC54-1-A
37	588	34.4	927	12	BG033451	BG033451 602301836
38	586.6	34.3	624	10	BE280681	BE280681 601155947
39	585.4	34.2	636	14	BM748747	BM748747 K-EST0023
40	584.6	34.2	842	9	AL580770	AL580770 AL580770
41	580.6	33.9	904	10	BE621185	BE621185 601494101
42	580	33.9	649	12	BF947767	BF947767 MR3-NN022
43	580	33.9	824	12	BG106653	BG106653 602290482
44	579.4	33.9	787	12	BG705505	BG705505 602688609
45	578	33.8	621	10	BE292913	BE292913 601105354

#### ALIGNMENTS

RESULT 1	AK012134	2612 bp	mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
LOCUS	AK012134	2612 bp	mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
DEFINITION	AK012134	2612 bp	mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
ACCESSION	AK012134	2612 bp	mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
VERSION	AK012134.1	GI:12848691	mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
KEYWORDS	HTC; CAP trapper.		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
SOURCE	mus musculus (strain:C57BL/6J) 10 days embryo cdna to mRNA, clone:2610524B01.		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
ORGANISM	mus musculus		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
AUTHORS	Carninci,P. and Hayashizaki,Y.		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
TITLE	High-efficiency full-length cdna cloning		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
MEDLINE	99279253		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
PUBMED	10349636		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
REFERENCE	1		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
AUTHORS	Carninci,P. and Hayashizaki,Y.		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cdna libraries for rapid discovery of new genes		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
MEDLINE	20499374		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002
PUBMED	11042159		mus musculus 10 days embryo whole body cdna, RIKEN full-length enriched library, clone:2610524B01:homolog to ANTIGEN ART1/P17, full insert sequence.	linear	HTC 19-JAN-2002

REFERENCE  
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multipillarary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G.,  
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsi, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Washima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Rindwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyo-Oka, K., Wang, K.H., Weitz, C., Wittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohtsuki, S.,  
and Hayashizaki, Y.

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE

## AUTHORS

## FEATURES

## source

## CDS

## REFERENCE







CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LLM13291 row: e column: 03  
High quality sequence stop: 653.  
Location/Qualifiers  
1. 865  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6046058"  
/clone\_lib="NIH\_MGC\_92"  
/tissue\_types="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
214 a 243 c 210 g 197 t 1 others

BASE COUNT 214 a 243 c 210 g 197 t 1 others  
ORIGIN

Query Match 48.2%; Score 824.4; DB 14; Length 865;  
Best Local Similarity 99.4%; Pred. No. 1.2e-140;  
Matches 858; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 370 ACCAATATCATCAAGCCAGACCAACAGAGTTCCTCGATTGCTCCACGGGAGTTCCG 429  
|||||  
Db 1 ACCAATATCATCAAGCCAGACCAACAGAGTTCCTCGATTGCTCCACGGGAGTTCCG 60  
|||||

QY 430 TCTGTGGAGTCCATGACCCACCCCTCGACCAACCTCAGCCAAACGCGGAAGCCCC 489  
|||||

Db 61 TCTGTGGAGTCCATGACCCACCCCTCGACCAACCTCAGCCAAACGCGGAAGCCCC 120  
|||||

QY 490 CACTATGCTGGACATCCCTCAGAGCATGTAGTCTCACCATCCATACAGTTCAGTTGAT 549  
|||||

Db 121 CACTATGCTGGACATCCCTCAGAGCATGTAGTCTCACCATCCATACAGTTCAGTTGAT 180  
|||||

QY 550 TCAGCACACCGAGCTCTTCGCAACCTTAFTGCCACAGCTCAGGCCCGCAAGATCAGCAGCA 609  
|||||

Db 181 TCAGCACACCGAGCTCTTCGCAACCTTAFTGCCACAGCTCAGGCCCGCAAGATCAGCAGCA 240  
|||||

QY 610 GACAGAAGGTGTAACAACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTGGTCCACCTCC 669  
|||||

Db 241 GACAGAAGGTGTAACAACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTGGTCCACCTCC 300  
|||||

QY 670 TCTCCCTGATGACCTCTGCTGTTAGATTGTAAGAATCCCAATGCACCAATCCAGATCCG 729  
|||||

Db 301 TCTCCCTGATGACCTCTGCTGTTAGATTGTAAGAATCCCAATGCACCAATCCAGATCCG 360  
|||||

QY 730 GCACACTGACCCAGAGAGTCACTTTATCTGTGGAAAGGGAGACCTGTGACTGAACCTCAG 789  
|||||

Db 361 GCACACTGACCCAGAGAGTCACTTTATCTGTGGAAAGGGAGACCTGTGACTGAACCTCAG 420  
|||||

QY 790 CTGGCACTCTGTGCGCAGTCTCTTACAGGAGTGGCCACAACTCTGCCCCACCGGG 849  
|||||

Db 421 CTGGCACTCTGTGCGCAGTCTCTTACAGGAGTGGCCACAACTCTGCCCCACCGGG 480  
|||||

QY 850 CTTTGACTGTGTAATAGAGTGTCTGAGACCCCTAACTGATGTGGCAGATGAGTATTG 909  
|||||

Db 481 CTTTGACTGTGTAATAGAGTGTCTGAGACCCCTAACTGATGTGGCAGATGAGTATTG 540  
|||||

QY 910 CCTTAAGTTTACCAAGTTGCTGGTTTGTGTGGACCGGGAGCGCGCTGGACAGAC 969  
|||||

Db 541 CCTTAAGTTTACCAAGTTGCTGGTTTGTGTGGACCGGGAGCGCGCTGGACAGAC 600  
|||||

QY 970 TCCTTTTCTGATGTGATGAGCAGGATTTCCATGAAGTGGGTATTGGCAGTGTGCTCTC 1029  
|||||

Db 601 TCCTTTTCTGATGTGATGAGCAGGATTTCCATGAAGTGGGTATTGGCAGTGTGCTCTC 660  
|||||

QY 1030 CTCCAGAAAGTTCTGCGACACCGGCATCAAGGACTATCAGATTACATGCTACAGATTAG 1089  
|||||

|||||  
Db 661 CTCCCAAGTCTTGCGACACCGCATCAAGGACTATCAGATTACATGCTACAGATTAG 720  
|||||

QY 1090 TAAGCAACTCTCTGAAGAATATGAAGGATTGTCAATCCTGTGAAGGCCACAGAGGACGC 1149  
|||||

Db 721 TAAGCAACTCTCTGGAGAATATGAAGGATTGTCAATCCTGTGAAGGCCACAGAGGACGC 780  
|||||

QY 1150 TAAACCTGTGAAGATCAA-GGAGGAACCTGTGAGCGACATCAC-TTTTCTGTGAGT-GA 1206  
|||||

Db 781 TAAACCTGTGAAGATCAAAGGGAGGAGCACTGTGAGCGACATCACTTTTCTGTGTCANTGA 840  
|||||

QY 1207 GGAGCTGGAGGCTGACCTTCTT 1229  
|||||

Db 841 GGAGCTGGAGGCTGACCTTCTT 863  
|||||

RESULT 5  
BM545988 1214 bp mRNA linear EST 20-FEB-2002  
LOCUS  
DEFINITION AGENCOURT\_6505156 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:5588387  
5', mRNA sequence.  
ACCESSION BM545988  
VERSION BM545988.1 GI:18778575  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1214)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Invitrogen  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LLM12359 row: g column: 12  
High quality sequence start: 135  
High quality sequence stop: 529.  
Location/Qualifiers  
1. 1214  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5588387"  
/clone\_lib="NIH\_MGC\_125"  
/lab\_host="DH10B"  
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;  
Site\_1: EcoRV (destroyed); Site\_2: NotI; RNA source pool  
of three ovaries, from females ranging in age from 38 to  
49 yo. Library is oligo-dT primed and directionally cloned  
(EcoRV site is destroyed upon cloning). Average insert  
size 2.1 kb, insert size range 1-3.5 kb. Library is  
normalized and enriched for full-length clones and was  
constructed by C. Gruber (Invitrogen). Research Genetics  
tracking code 036."  
268 a 347 g 312 g 250 t 37 others

BASE COUNT 268 a 347 g 312 g 250 t 37 others  
ORIGIN

Query Match 47.8%; Score 817.4; DB 13; Length 1214;  
Best Local Similarity 97.4%; Pred. No. 2e-139;  
Matches 827; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 ACAGCATCTTGGCTCAGGCTCTCGAGGTCCAGACAGCGCCGCGCTCTGCGACG 60  
|||||

Db 116 ACCCNCACCTTCCCTCANGCTCTCNAGGTCCAAACAGCGCCGCGCTCTGCGACG 175  
|||||

QY 61 CAGCAGTAATAGTGTGGTACCTCTTGTCTCGTTCAGGTCCAGACCTCCCGCTTCC 120  
|||||

Db 176 CNGCAGTGAATAGTGTGGTACCTCCTGTCTCGGTTCAGGTCCAGACCTCCCGCTTCC 235  
QY 121 GGCTGCTTGAACGTCAGCGGACCTCAGGACCTGTGATTGGCGCTGGCGGGGAC 180  
Db 236 GGCTGCTTGAACGTCAGCGGACCTCAGGACCTGTGATTGGCGCTGGCGGGGAC 295  
QY 181 GTGACCGAGAAACCCCTGGAGGACTTGGGCAATTCCTTTGGGCTCCGTGCTTCTTCG 240  
Db 296 GTGACCGAGAAACCCCTGGAGGACTTGGGCAATTCCTTTGGGCTCCGTGCTTCTTCG 355  
QY 241 TCCTCCTTCGGGCAAGGATCTACATATATAGTCTTTGACCGACACAGATGCTCGCA 300  
Db 356 TGCTCCTTCGGGCAAGGATCTACATATATAGTCTTTGACCGACACAGATGCTCGCA 415  
QY 301 TTTGATAAATGTTTGTGAACCTTGAAGACATATGGCAATGAATCTGCAAGATACGT 360  
Db 416 TTTGATAAATGTTTGTGAACCTTGAAGACATATGGCAATGAATCTGCAAGATACGT 475  
QY 361 GGGAGATACCAATATATCAATCAAGCCAGACCAAGAGTTCTTTCGATTGCTCCACG 420  
Db 476 GGGAGATACCAATATATCAATCAAGCCAGACCAAGAGTTCTTTCGATTGCTCCACG 535  
QY 421 GGAGTTCCGCTGTGGTGAAGTCCATGACCCCTGACCAACCTCAGCCCAAGGCC 480  
Db 536 GGAGTTCCGCTGTGGTGAAGTCCATGACCCCTGACCAACCTCAGCCCAAGGCC 595  
QY 481 GAAGCCCCCCTACTGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGAT 540  
Db 596 GAAGCCCCCCTACTGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGAT 655  
QY 541 TCAGTTGATTCAGCACACACCGACGCTTTCGCAACCTTATGCCCAGCTCAGGCCAGAA 600  
Db 656 TCAGTTGATTCAGCACACACCGACGCTTTCGCAACCTTATGCCCAGCTCAGGCCAGAA 715  
QY 601 TCAGCAGCAGACAGAGAGTGTAAACCTGAGAGAGTGAACCTCTTCCCTGCGCCCTGG 660  
Db 716 TCAGCAGCAGACAGAGAGTGTAAACCTGAGAGAGTGAACCTCTTCCCTGCGCCCTGG 775  
QY 661 CTCACCTCCTCTCCCTGATGACCTCCTGCTTTAGATTGTAAAGATCCCAATGCAACCAT 720  
Db 776 CTCACCTCCTCTCCCTGATGACCTCCTGCTTTAGATTGTAAAGATCCCAATGCAACCAT 835  
QY 721 CCAGATCCGGCAGTACCCAGAGAGTACTTTTATCGTGGGAAAGGAGACCTGTGAC 780  
Db 836 CCAGATCCGGCAGTACCCAGAGAGTACTTTTATCGTGGGAAAGGAGACCTGTGAC 895  
QY 781 TGAACCTCAGCTGGCACTCCTGTGCGCAGCTCCTTACCAGCAGTGGCCCAATCCTGCG 840  
Db 896 TGAACCTCAGCTGGCACTCCTGTGCGCAGCTCCTTACCAGCAGTGGCCCAATCCTGCG 955  
QY 841 CCACCCGGG 849  
Db 956 CCCACCCGG 964

RESULT 6  
BQ924589  
LOCUS  
DEFINITION BQ924589 923 bp mRNA linear EST 20-AUG-2002  
5' mRNA sequence.  
ACCESSION BQ924589  
VERSION BQ924589.1 GI:22339620  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 923)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP/Gazdar  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM2600 row: c column: 19  
High quality sequence stop: 635.  
location/Qualifiers  
1. 923  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:642730"  
/clone\_lib="NIH\_MGC\_18"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."  
BASE COUNT 209 a 283 c 221 g 210 t  
ORIGIN  
Query Match 46.6%; Score 797.4; DB 14; Length 923;  
Best Local Similarity 96.7%; Pred. No. 9.6e-136;  
Matches 889; Conservative 0; Mismatches 21; Indels 9; Gaps 7;  
QY 27 AGGTCCACACAGCCGCCAGCCGCTCTGCGACGACGAGTGAATAGTGTGACTCTCT 86  
Db 1 AGGTCCACACAGCCGCCAGCCGCTCTGCGACGACGAGTGAATAGTGTGACTCTCT 60  
QY 87 TGTCTCGTTTCAGGTTCACAGACTTCCCGCTCTTCCGGCTGCCCTGAACTGACGGACCTC 146  
Db 61 TGTCTCGTTTCAGGTTCACAGACTTCCCGCTCTTCCGGCTGCCCTGAACTGACGGACCTC 120  
QY 147 AGGACCTGTATTTGGCGCTTCCCGCGGACCGCTGACCGAGGAAACCCCTGGAGGAC 206  
Db 121 AGGACCTGTATTTGGCGCTTCCCGCGGACCGCTGACCGAGGAAACCCCTGGAGGAC 180  
QY 207 TTGGCATTCTTGGGCTCCGCTGCTTCTTCTGCTGCTTCTTGGGCAAGGATCTCACA 266  
Db 181 TTGGCATTCTTGGGCTCCGCTGCTTCTTCTGCTGCTTCTTGGGCAAGGATCTCACA 240  
QY 267 TTATCAGTCTTTGACCGACACAGATGCTGCAATTTGATAAATGTTTGTGAACCTGAA 326  
Db 241 TTATCAGTCTTTGACCGACACAGATGCTGCAATTTGATAAATGTTTGTGAACCTGAA 300  
QY 327 GAGACATATGACATGATGATCTGCAAGATCTGGGAGAGATACCAATATCATCAAGCC 386  
Db 301 GAGACATATGACATGATGATCTGCAAGATCTGGGAGAGATACCAATATCATCAAGCC 360  
QY 387 AGACCAACAGAAAGTCTTTCGATTGCTCCACGCGGAGTTCGCTGCTGCTGGAAGTCCATG 446  
Db 361 AGACCAACAGAAAGTCTTTCGATTGCTCCACGCGGAGTTCGCTGCTGGAAGTCCATG 420  
QY 447 ACCACCCCTGCACCAACCCCTCAGCCCAACAGCCGAGCCGCCCACTATGCTGGACATCC 506  
Db 421 ACCACCCCTGCACCAACCCCTCAGCCCAACAGCCGAGCCGCCCACTATGCTGGACATCC 480  
QY 507 CCTCAGAGCCATGATGATCTCACCATCCATAGATTTCAGTTCAGTTCAGACACACCGACCTC 566  
Db 481 CCTCAGAGCCATGATGATCTCACCATCCATAGATTTCAGTTCAGTTCAGACACACCGACCTC 540  
QY 567 TTCGCAACCTTATTTGCCACAGCTCAGGCCCGGAGATTCAGCAGCAGACAGAGGTGTAAAA 626  
Db 541 TTCGCAACCTTATTTGCCACAGCTCAGGCCCGGAGATTCAGCAGCAGACAGAGGTGTAAAA 600  
QY 627 CTGAAGAGAGTGAACCTCTTCCCTGCTGTCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 685

FEATURES  
source







QY 1016 GCAGTGTGCTCTCCCTCCAGAAAGTTCTGGGAGCACCAGCATCAAGGACTATCACAGTTAC 1075  
 Db 659 -GGCAGGTGCTCTCCCTCCAGAAAGTTCTGGGAGCACCAGCATCAAGGACTATCACAGTTAC 717  
 QY 1076 ATGCTACAGATTAGTAAGCAACTCTCTGAAGA- --ATATGAAGGATTGTCAATCCTG- - 1130  
 Db 718 ATGCTACAGATTAGTAAGCAACTCTCTGAAGACATATGAGAGGATGGTCAATCCTGGA 777  
 QY 1131 AGAAGGCCACAGAGGACGCTAAA-CCTGTGAAGATCAAGGAGGAACCTGTG-AGCGACAT 1188  
 Db 778 GGAAGGCCACAGAGGACGCTAAACCTCTGTGAAGATCAAGGAGGAACCTGTGAAGGACAT 837  
 QY 1189 CACTTTTCTGTGCTAGT-AGGAGCTGGAGGC-TGACCTTGTCTTGGAGACCAAGTCACTG 1246  
 Db 838 CACGTTTACCTGCAGGAGGAGCTGGAGGCATGACCTGCACTCGGGAGACCAAGTGCCT 897  
 QY 1247 CCTATGGGAGTCTTGGGCG 1266  
 Db 898 GCCTAGAGAGAGCAGGGGCG 917

## RESULT 11

## B1770385

LOCUS 60305622F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5205245 5',  
 mRNA sequence.

ACCESSION B1770385

VERSION B1770385.1 GI:15761963

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 771)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11514 row: k column: 06

High quality sequence stop: 768.

Location/Qualifiers

1..771

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5205245"

/clone\_lib="NIH\_MGC\_122"

/lab\_host="DH10B"

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;

Site 1: NotI; Site 2: EcoRV (destroyed); RNA source

anonymous pool of 24 week female lung, 16 week female

spleen, and 20-22 week male spleens. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.4 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 026. Note:

this is a NIH\_MGC Library."

BASE COUNT 175 a 243 c 184 g 168 t 1 others

## ORIGIN

Query Match

Best Local Similarity 42.2%; Score 721.4; DB 13; Length 771;

Matches 737; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 8 CTTTGCCTCAGCCCTCTCGAGTCCAGACAGCCGCCAGCCGCTCTCGGAGCGCAGCAGT 67  
 Db 26 CTTTGCCTCAGCCCTCTCGAGTCCAGACAGCCGCCAGCCGCTCTCGGAGCGCAGCAGT 85  
 QY 68 GAATAGTGTGTACCTCTCTGTCGGTTTCAAGTCCAGACCTCCGCCGCTCTTCGGGCTGCC 127  
 Db 86 GAATAGTGTGTACCTCTCTGTCGGTTTCAAGTCCAGACCTCCGCCGCTCTTCGGGCTGCC 145  
 QY 128 CTGAACGTCAGCGCACCTTCAGGACCTCTGTGATTTGCGCTCTGCGCGCGGCGGACCGT 187  
 Db 146 CTGAACGTCAGCGCACCTTCAGGACCTCTGTGATTTGCGCTCTGCGCGCGGCGGACCGT 205  
 QY 188 AGGAAACCCCTGGAGGACCTTGGGCATTCCTTTGGGCTCCGTCGCTCTTCTTGGTCTCT 247  
 Db 206 AGGAAACCCCTGGAGGACCTTGGGCATTCCTTTGGGCTCCGTCGCTCTTCTTGGTCTCT 265  
 QY 248 TTCGGGCAAGGATCTCACATTTATGATTTTACCGACACAGAAATGCTTGGCATTTGATA 307  
 Db 266 TTCGGGCAAGGATCTCACATTTATGATTTTACCGACACAGAAATGCTTGGCATTTGATA 325  
 QY 308 AATGTTTGTGAACCTTGAAGAGACATATGGACAATGAAATCTGAAAAGATCTTGGGAGAG 367  
 Db 326 AATGTTTGTGAACCTTGAAGAGACATATGGACAATGAAATCTGAAAAGATCTTGGGAGAG 385  
 QY 368 ATACCAATATCATCAAGCCAGCACCACCAAGAAAGTTTCTTTCGATTTGCTCCCGAGGATTC 427  
 Db 386 ATACCAATATCATCAAGCCAGCACCACCAAGAAAGTTTCTTTCGATTTGCTCCCGAGGATTC 445  
 QY 428 CGTCTGGTGAAGTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 487  
 Db 446 CGTCTGGTGAAGTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 505  
 QY 488 CCCACTATGCTGGACATCCCTCTCAGAGCATGATGATGATGATGATGATGATGATGATGAT 547  
 Db 506 CCCACTATGCTGGACATCCCTCTCAGAGCATGATGATGATGATGATGATGATGATGATGAT 565  
 QY 548 ATTCAGCACAACCGACGCTTTCGCAACCTTATGTCACAGCTCAGGCCCCAGAAATCAGAG 607  
 Db 556 ATTCAGCACAACCGACGCTTTCGCAACCTTATGTCACAGCTCAGGCCCCAGAAATCAGAG 625  
 QY 608 CAGACAGAAGTG- -TAAAACTGAAGAGTGAACCTTCCCTCGTGGCCCTGGTGCAC 665  
 Db 626 CAGACAGAAGTGTTAAAACTGAAGAGTGAACCTTCCCTCGTGGCCCTGGTGCAC 685  
 QY 666 CTCCTCTCCCTGATGACCTCTGCTTGTAGATTGTAGAAATCCCAATCCCAATCCCAATCC 725  
 Db 686 CTCCTCTCCCTGATGACCTCTGCTTGTAGATTGTAGAAATCCCAATCCCAATCCCAATCC 745  
 QY 726 TCCGGCAGAGTGACCCAGAGAGTGA 750  
 Db 746 TCCGGCAGAGTGACCCAGAGAGTGA 770

## RESULT 12

## BQ576180/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

BQ576180 735 bp mRNA linear EST 19-JUN-2002  
 UI-H-EZ1-bbi-i-21-0-UI.s1 NCI\_CGAP\_Ch2 Homo sapiens cDNA clone  
 UI-H-EZ1-bbi-i-21-0-UI 3', mRNA sequence.

BQ576180

BQ576180.1 GI:21479497

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 735)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of

Orthopedics  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

Location/Qualifiers

source

1..735  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-H-EZ1-bb1-1-21-0-UI"  
/clone\_lib="NCI\_CGAP\_Ch2"  
/tissue\_type="Chondrosarcoma Grade II"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)  
with a modified polylinker; Site.1: EcoR I; Site.2: Not I;  
NCI\_CGAP\_Ch2 is a normalized cDNA library containing the  
following tissue(s): Chondrosarcoma Grade II. The library  
was constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pT73-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
TGATCAGCT.

TAG LIB=UI-H-EZ1

TAG TISSUE=grade-2-chondrosarcoma

TAG\_SEQ=ATCTAATATG

177 a 187 c 158 g 211 t 2 others

BASE COUNT 177 a 187 c 158 g 211 t 2 others

ORIGIN

Query Match 41.5%; Score 710.8; DB 14; Length 735;

Best Local Similarity 98.6%; Pred. No. 6.4e-120;

Matches 726; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

964 ACAGACTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

965 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

966 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

967 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

968 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

969 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

970 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

971 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

972 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

973 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

974 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

975 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

976 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

977 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

978 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

979 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

980 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

981 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

982 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

983 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

984 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

985 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

986 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

987 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

988 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

989 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

990 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

991 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

992 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

993 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

994 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

995 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

996 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

997 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

998 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

999 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1000 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1001 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1002 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1003 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1004 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1005 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1006 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1007 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1008 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1009 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1010 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1011 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1012 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1013 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1014 ACATAGTCTTTCTGATGTGATGAGGAGGATTCATGAAGTGGTATGGCAGTGT 1023

1444 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1445 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1446 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1447 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1448 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1449 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1450 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1451 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1452 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1453 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1454 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1455 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1456 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1457 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1458 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1459 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1460 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1461 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1462 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1463 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1464 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1465 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1466 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1467 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1468 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1469 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1470 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1471 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1472 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1473 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1474 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1475 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1476 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1477 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1478 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1479 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1480 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1481 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1482 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1483 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1484 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1485 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1486 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1487 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1488 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1489 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1490 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1491 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1492 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1493 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1494 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1495 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1496 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1497 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1498 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1499 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1500 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1501 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1502 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1503 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1504 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1505 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1506 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1507 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1508 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1509 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1510 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1511 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1512 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1513 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1514 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1515 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1516 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1517 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503

1518 TATGTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCTGATCACTAGATAGCAGCTA 1503



```
Db 2 GAGGTCCAGACAGCGCCGCCGCTCTGCGACGACGAGTGAATAGTGTGTAACCTCC 61
QY 86 TTGTCCTCGGTCAGGTCCAGACCTCCCGCTCTCCGGCTGCCCTGAAGCTCAGGCGACCT 145
Db 62 TTGTCCTCGGTCAGGTTCAGACCTCCCGCTCTCCGGCTGCCCTGAAGCTCAGGCGACCT 121
QY 146 CAGGACCTGTGATTGGCGCTCGCGCGCGGACCGTGCACGAGSAAACCCCTCGAGGGA 205
Db 122 CAGGACCTGTGATTGGCGCTCGCGCGCGGACCGTGCACGAGSAAACCCCTCGAGGGA 181
QY 206 CTTGGGCAATTCCTGGGCTCGGTCTTCTGCTGCTCTTTCGGGCAAGATCTCAC 265
Db 182 CTTGGGCTTCTTGGGCTCGGTCTTCTGCTGCTCTTTCGGGCAAGATCTCAC 241
QY 266 ATTATCAGTCTTTGACCCACACAGATGCCTGGCATTTGATAATGTTTGTGAACCTGA 325
Db 242 ATTATCAGTCTTTGACCCACACAGATGCCTGGCATTTGATAATGTTTGTGAACCTGA 301
QY 326 AGAGACATATGACAAATGATCTGCAAGATACCTGGGAGAGATACCAATATCATCAAGC 385
Db 302 AGAGACATATGACAAATGATCTGCAAGATACCTGGGAGAGATACCAATATCATCAAGC 361
QY 386 CAGACCAACAGAGTTCCTTCGATTGCTCCACGGGAGTTCCGCTGCTGGTGAAGTCCAT 445
Db 362 CAGACCAACAGAGTTCCTTCGATTGCTCCACGGGAGTTCCGCTGCTGGTGAAGTCCAT 421
QY 446 GACCCACCCCTGC-ACCAACCCCTCAGCCAACGCGAAGCCGCCCTACCTATGCTGGACAT 504
Db 422 GAACCGACCTTCGACCAAGCTTCAGGCAACAGCCGGAAGCCGCCCTACCTATGCTGGACAT 481
QY 505 CCCCTCAGAGGCAATGATGCTACATCCAT-AGATTTCAGTTGATTTCAGCAACACCGAC 563
Db 482 CCCCTCAGAGGCAATGATGCTACATCCATCATAAAGATTTCAGTTTCAGCAACACCGAC 541
QY 564 GTCCTCGAAGCTTATTCGACAGCTCAGGCCCGCAAGATCAGCAGCAGACAGAAGGTGTA 623
Db 542 GTCCTCGAAGCTTATTCGACAGCTCAGGCCCGCAAGATCAGCAGCAGACAGAAGGTGTA 601
QY 624 AAACCTGAAGAGAGTGAACCTCTCCCTCGTCCCTGGGTCACTCCCTCCTGATGACC 683
Db 602 AAACCTGAAGAGAGTGAACCTCTCCCTCGTCCCTGGGTCA-CTCCCTCCTGATGACC 660
QY 684 TCCGCTCTTATGATTGAAGATCCCAATGACCATTCAGATCCGCGACAGTCAACCCAG 743
Db 661 TCATGCTCTTATGATTGAAGATCCCAATGACCATTCAGATCCGCGACAGTCAACCCAG 720
QY 744 AGAGTGACTTTATCGTGGGAAGGGAAGCTGTGACTGACTGACTGACTGACTGACTGCTC 803
Db 721 AGGAGTGACTATATCGTGGGAAGGGAAGCTGTGACTGACTGACTGACTGACTGACTGTA 780
QY 804 GG-CAGCTCCCTCTACCGAGCAGTGGCCACA 832
Db 781 GGACAATACCTCTACCGAGCAGTGGACAAA 810

RESULT 14
LOCUS BM728984
DEFINITION UI-E-E01-aiw-c-15-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone
AUTHORS UI-E-E01-aiw-c-15-0-UI 5', mRNA sequence.
ACCESSION BM728984
VERSION BM728984.1 GI:19050317
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 718)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
```

## COMMENT

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Tissue: Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

## FEATURES

## source

Location/Qualifiers  
1..718  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-E01-aiw-c-15-0-UI"  
/clone\_lib="UI-E-E01"  
/tissue\_type="fetal eye"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-E-E01 is a normalized cDNA library containing the  
following tissue(s): fetal eye. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pT73-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
cgggtatacc. This library was created for the program, Gene  
Discovery in the Visual System, supported by National Eye  
Institute (NEI)."

BASE COUNT 170 a 207 c 172 g 168 t 1 others  
ORIGIN

Query Match 41.2%; Score 704.4; DB 14; Length 718;  
Best Local Similarity 99.6%; Pred. No. 9.4e-119;  
Matches 716; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 395 AGAGTTTCCTTCGATTGCTCCACGGAGTTCGCTGTTGGGAAGTCCATGACCCACCC 454  
Db 1 AGAAGTTCCCTTCGATTGCTCCACGGAGTTCGCTGTTGGGAAGTCCATGACCCACCC 60  
QY 455 CTGCACCAACCCCTCAGCCACAAGCCGAAGCCGCCCTACTATGCTGGACATCCCTTCAGAG 514  
Db 61 CTGCACCAACCCCTCAGCCACAAGCCGAAGCCGCCCTACTATGCTGGACATCCCTTCAGAG 120  
QY 515 CCATGTAGTCTCACCATTCCATACGATTTCAGTTGATTTCAGCACACACCGCTCTTCGCAAC 574  
Db 121 CCATGTAGTCTCACCATTCCATACGATTTCAGTTGATTTCAGCACACACCGCTCTTCGCAAC 180  
QY 575 CTTATTGCCACAGCTCAGGCCCAAGTTCAGCAGCAGACAGAAAGTGTAAAACTGAAGAG 634  
Db 181 CTTATTGCCACAGCTCAGGCCCAAGTTCAGCAGCAGACAGAAAGTGTAAAACTGAAGAG 240  
QY 635 AGTGAACCTCTTCCCTCGTCCCTGGTGGTCACTCCTCTCCCTGATGACCTCCTCCCTTTA 694  
Db 241 AGTGAACCTCTTCCCTCGTCCCTGGTGGTCACTCCTCTCCCTGATGACCTCCTCCCTTTA 300  
QY 695 GATTGTAGATTCATTCAGTCCATTCAGTCCAGATCCGGCACAGTACCACAGAGATGACTTT 754  
Db 301 GATTGTAGATTCATTCAGTCCATTCAGTCCAGATCCGGCACAGTACCACAGAGATGACTTT 360  
QY 755 TATCGTGGGAAAGGGGAACCTGTGACTGAACTCAGCTGGCACTCCTCTGTGGCAGCTCCTC 814  
Db 361 TATCGTGGGAAAGGGGAACCTGTGACTGAACTCAGCTGGCACTCCTCTGTGGCAGCTCCTC 420





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 18, 2003, 09:48:28 ; Search time 0.872038 seconds  
(without alignments)  
881.929 Million cell updates/sec

Title: US-09-857-308-5  
Perfect score: 37  
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	89.2	426	2 S51016	sodium-translocati
2	32	86.5	865	2 T41685	probable gamma-ada
3	31	83.8	269	2 T31318	hypothetical prote
4	30	81.1	307	2 S55596	hypothetical prote
5	30	81.1	312	2 H71948	probable tetraacyl
6	30	81.1	410	2 H81918	probable sodium-tr
7	30	81.1	410	2 C81185	Na(+)-translocatin
8	30	81.1	413	2 AF0393	NADH2 dehydrogenas
9	29	78.4	91	2 B82893	hypothetical prote
10	29	78.4	143	2 F82292	hypothetical prote
11	29	78.4	233	2 S77776	triacylglycerol li
12	29	78.4	264	2 J41111	triacylglycerol li
13	29	78.4	299	2 F83795	ABC transporter (A
14	29	78.4	363	2 A81837	hypothetical prote
15	29	78.4	370	2 C71926	cag island protein
16	29	78.4	421	2 S26246	glutamate/aspartat
17	29	78.4	421	2 S26247	glutamate/aspartat
18	29	78.4	502	2 T16921	hypothetical prote
19	29	78.4	503	2 J05078	glutamate transpor
20	29	78.4	507	2 T34005	hypothetical prote
21	29	78.4	518	2 T29633	hypothetical prote
22	29	78.4	532	2 T23481	hypothetical prote
23	29	78.4	565	2 I38399	glutamate/aspartat
24	29	78.4	572	2 A55676	excitatory amino a
25	29	78.4	573	2 S28901	glutamate transpor
26	29	78.4	574	2 I37426	glutamate transpor
27	29	78.4	574	2 J04262	glutamate transpor
28	29	78.4	574	2 I38432	excitatory amino a
29	29	78.4	575	2 T29354	hypothetical prote

two-component sens  
semaphorin B - mou  
hypothetical prote  
sensor protein evg  
probable sensor fo  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypotHet  
hypothetical prote  
hypothetical 22K p  
probable shikimate  
probable phenylace  
germin-like protei  
hemagglutinin - Ma

ALIGNMENTS

RESULT 1

S51016

sodium-translocating NADH dehydrogenase (ubiquinone) (EC 1.6.5.-) nqrB chain - Vibrio  
N:Alternate names: NADH-ubiquinone oxidoreductase nqrB protein  
C:Species: Vibrio alginolyticus  
C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 21-Jul-2000  
C:Accession: S51016  
R:Beattie, P.; Tan, K.; Bourne, R.M.; Leach, D.; Rich, P.R.; Ward, F.B.  
FEBS Lett. 356, 333-338, 1994  
A:Title: Cloning and sequencing of four structural genes for the Na(+)-translocating  
A:Reference number: S51013; MUID:95104445; PMID:7805867  
A:Accession: S51016  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-426 <BEA>  
A:Cross-references: EMBL:Z37111; NID:g663268; PIDN:CAA85477.1; PID:g663271  
C:Genetics:  
A:Gene: nqrB  
C:Keywords: NAD; oxidoreductase

Query Match 89.2%; Score 33; DB 2; Length 426;  
Best Local Similarity 75.0%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
DB 38 LYEA VATV 45  
||:||||:

RESULT 2

T41685

probable gamma-adaptin - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T41685  
R:Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z21742  
A:Accession: T41685  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-865 <WED>  
A:Cross-references: EMBL:AL117183; PIDN:CAB54865.1; GSPDB:GN00068; SPDB:SPCP1E11.06  
A:Experimental source: strain 972h-; clone pl p1E11  
C:Genetics:  
A:Gene: SPDB:SPCP1E11.06  
A:Map position: 3

Query Match 86.5%; Score 32; DB 2; Length 865;  
Best Local Similarity 87.5%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
| | | | |  
Db 317 LYQAVRTI 324

### RESULT 3

T31318  
hypothetical protein - Cenarchaeum symbiosum  
C:Species: Cenarchaeum symbiosum  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C:Accession: T31318  
R:Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.  
J. Bacteriol. 180, 5003-5009, 1998  
A:Title: Genomic analysis reveals chromosomal variation in natural populations of the un  
A:Reference number: Z20994; MUID:98422450; PMID:9748430  
A:Accession: T31318  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-269 <SCH>  
A:Cross-references: EMBL:AF083072; NID:q3599393; PID:q3599404; PIDN:AAC62709.1

```
Query Match      83.8%; Score 31; DB 2; Length 269;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 LYQAVATI 8  
Db 174 LYQAVPTV 181

## RESULT 4

RESOUR 4  
 S55596  
 hypothetical protein E2 - equine herpesvirus 2  
 C:Species: equine herpesvirus 2  
 C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999  
 C:Accession: S55596  
 P:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.  
 J. Mol. Biol. 249, 520-528, 1995  
 A:Title: The DNA sequence of equine herpesvirus 2.  
 A:Reference number: S55594; MUID:7783207  
 A:Accession: S55596  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1307 <TEL>  
 A:Cross-references: GB:U00824; NID:g695172; PIDN:AAIC3789.1; PID:g695174  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, Feb.

Query Match 81.1%; Score 30; DB 2; Length 307;  
Best Local Similarity 85.7%; Pred. No. 40;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YQAVATI 8  
Db 213 YQAVSTI 219

## RESULT 5

H71948  
 probable tetraacyldisaccharide 4'-kinase (EC 2.7.1.130) lpK [similarity] - Helicobacter  
 C.Species: Helicobacter pylori  
 A.Variety: strain J99  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Mar-2001  
 C:Accession: H71948  
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A:Reference number: A71800; MUID:99120557; PMID:9923682  
 A:Accession: H71948  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-312 <ARN>  
 A:Cross-references: GB:AE001467; GB:AE001439; NID:c4154826; PIDN:AAD05884.1; PID:c415483

Db 26 LYEAAATI 33

RESULT 8

AF0393

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain B [imported] - Yersinia pestis (strain AF0393)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 03-Jun-2002  
C:Accession: AF0393  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; et al.; M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0393  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92474.1; PID:g15981175; GSPDB:GN00175  
C:Genetics:  
A:Gene: nqrB  
C:Keywords: oxidoreductase

Query Match 81.1%; Score 30; DB 2; Length 413;  
Best Local Similarity 75.0%; Pred. No. 54;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

|||||

Db 26 LYEAAATI 33

RESULT 9

B82893

hypothetical protein UUA424.1 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: B82893  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H. submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min  
A:Reference number: A82870  
A:Accession: B82893  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <GLA>  
A:Cross-references: GB:AE002139; GB:AF222894; NID:96899405; PIDN:AAF30836.1; GSPDB:GN001  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: UUA424.1  
A:Genetic code: SGC3

Query Match 78.4%; Score 29; DB 2; Length 91;  
Best Local Similarity 87.5%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

|||||

Db 53 LLQAVATI 70

RESULT 10

F82292

hypothetical protein VC0697 [imported] - Vibrio cholerae (strain N15961 serogroup O1)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: F82292  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; et al.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P. l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: F82292

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-143 <HEU>

A:Cross-references: GB:AE004155; GB:AE003852; NID:9655127; PIDN:AAF93862.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N15961; biotype El Tor

C:Genetics:

A:Gene: VC0697

A:Map position: 1

Query Match 78.4%; Score 29; DB 2; Length 143;

Best Local Similarity 75.0%; Pred. No. 30;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

|||||

Db 31 LYQAPATL 38

RESULT 11

S77776

triacylglycerol lipase (EC 3.1.1.3) - Mycoplasma capricolum (fragment)

N:Alternate names: Lipase; protein MC073

C:Species: Mycoplasma capricolum

C:Date: 09-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 21-Jul-2000

C:Accession: S77776

R:Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert, Mol. Microbiol. 16, 955-967, 1995

A:Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its phys

A:Reference number: S77739; MUID:96059641; PMID:7476192

A:Accession: S77776

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <BOR>

A:Cross-references: EMBL:Z33059; NID:9531610; PIDN:CAA83733.1; PID:94379137

A:Experimental source: AFCC 27343

C:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1994

C:Comment: This enzyme, serine esterase enzyme, is widely distributed throughout an

in fatty acids, and hydrolyses ester bonds of triacylglycerols to yield free fatty ac

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: triacylglycerol lipase 1

C:Keywords: carboxylic ester hydrolase

Query Match 78.4%; Score 29; DB 2; Length 233;

Best Local Similarity 62.5%; Pred. No. 50;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8

|||||

Db 165 IYQAIKTI 172

RESULT 12

JC4111

triacylglycerol lipase (EC 3.1.1.3) 3 - Mycoplasma mycoides subsp. mycoides

N:Alternate names: lipase

C:Species: Mycoplasma mycoides subsp. mycoides

C:Date: 23-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 07-Dec-1999

C:Accession: JC4111

R:Rawadi, G.; Lalanne, J.L.; Roulland-Dussoix, D.

Gene 158, 107-111, 1995

A:Title: Cloning and characterization of the lipase operon from Mycoplasma mycoides s

A:Reference number: JC4109; MUID:95309706; PMID:7789792

A:Accession: JC4111

A:Molecule type: DNA

A:Residues: 1-264 <RAW>

A:Cross-references: GB:U17036

C:Comment: This enzyme, serine esterase enzyme, is widely distributed throughout an

in fatty acids, and hydrolyses ester bonds of triacylglycerols to yield free fatty ac

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: triacylglycerol lipase 1  
C:Keywords: carboxylic ester hydrolase  
F:90-94/Region: conserved site

Query Match 78.4%; Score 29; DB 2; Length 264;  
Best Local Similarity 62.5%; Pred. No. 58;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
:|||: ||  
Db 197 IYQAIKTI 204

## RESULT 13

ABC transporter (ATP-binding protein) nata [imported] - Bacillus halodurans (strain C-12)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: F83795  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11059132  
A:Accession: F83795

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-299 <STO>  
A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BAB04885.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: nata

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 78.4%; Score 29; DB 2; Length 299;  
Best Local Similarity 85.7%; Pred. No. 66;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVAT 7  
:|||||  
Db 266 LFQAVAT 272

## RESULT 14

AEI837  
hypothetical protein all0245 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AEI837  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AEI837  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-363 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB77769.1; PID:g17135223; GSPDB:GN00179  
A:Experimental source: strain PCC 7120

C:Genetics:  
A:Gene: all0245

Query Match 78.4%; Score 29; DB 2; Length 363;  
Best Local Similarity 85.7%; Pred. No. 81;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YQAVATI 8  
||| |||  
Db 120 YQATATI 126

## RESULT 15

C71926  
cag island protein - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: C71926  
R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F

Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: C71926  
A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-370 <ARN>

A:Cross-references: GB:AE001482; GB:AE001439; NID:g4155018; PIDN:AAD06062.1; PID:g415

A:Experimental source: strain J99

C:Genetics:  
A:Gene: cagH

Query Match 78.4%; Score 29; DB 2; Length 370;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YQAVAT 7  
|||||  
Db 241 YQAVAT 246

Search completed: July 18, 2003, 10:00:41  
Job time : 2.87204 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:39:42 ; Search time 4490 Seconds  
(without alignments)  
11090.178 Million cell updates/sec

Title: US-09-857-308-2

Perfect score: 1711

Sequence: 1 acgcgatcttgcctcagc.....aaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_ba:\*\*

2: gb\_htg:\*\*

3: gb\_in:\*\*

4: gb\_om:\*\*

5: gb\_ov:\*\*

6: gb\_pat:\*\*

7: gb\_ph:\*\*

8: gb\_pl:\*\*

9: gb\_pr:\*\*

10: gb\_ro:\*\*

11: gb\_sts:\*\*

12: gb\_sy:\*\*

13: gb\_un:\*\*

14: gb\_vi:\*\*

15: em\_ba:\*\*

16: em\_fun:\*\*

17: em\_hum:\*\*

18: em\_in:\*\*

19: em\_mu:\*\*

20: em\_om:\*\*

21: em\_or:\*\*

22: em\_ov:\*\*

23: em\_pat:\*\*

24: em\_ph:\*\*

25: em\_pl:\*\*

26: em\_ro:\*\*

27: em\_sts:\*\*

28: em\_un:\*\*

29: em\_vi:\*\*

30: em\_htg\_hum:\*\*

31: em\_htg\_inv:\*\*

32: em\_htg\_other:\*\*

33: em\_htg\_mus:\*\*

34: em\_htg\_pln:\*\*

35: em\_htg\_rtd:\*\*

36: em\_htg\_mam:\*\*

37: em\_htg\_vrt:\*\*

38: em\_sy:\*\*

39: em\_htgo\_hum:\*\*

40: em\_htgo\_mus:\*\*

41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1691.8	98.9	2021	9	AF197954	AF197954 Homo sapi
2	1690.2	98.8	4261	9	AB018307	AB018307 Homo sapi
3	1688.6	98.7	2045	9	AK000329	AK000329 Homo sapi
4	1457.8	85.2	1481	9	AK000616	AK000616 Homo sapi
5	1340.8	78.4	2869	9	AF224759	AF224759 Homo sapi
6	1057	61.8	2616	10	BC031447	BC031447 Mus muscu
7	850.8	49.7	1046	6	AX013061	AX013061 Sequence
8	408	23.8	204153	9	AC074091	AC074091 Homo sapi
9	359	21.0	386	11	G25726	G25726 human STR E
10	356.8	20.9	173988	2	AC105160	AC105160 Mus muscu
11	356.8	20.9	232648	2	AC113276	AC113276 Mus muscu
12	352	20.6	155250	2	AC127781	AC127781 Rattus no
13	266.6	15.6	145971	2	AC096698	AC096698 Rattus no
14	253.2	14.8	145971	2	AC096698	AC096698 Rattus no
15	214.6	12.5	296	9	HS33F3F	HS33F3F
16	208.4	12.2	338	9	HS18D8R	HS18D8R
17	109	6.4	2954	9	AK001486	AK001486
18	83.4	4.9	2949	9	AY028435	AY028435 Homo sapi
19	53	3.1	125020	9	AF429315	AF429315 Homo sapi
20	49.2	2.9	154609	2	AC118085	AC118085 Rattus no
21	48.8	2.9	155250	2	AC127781	AC127781 Rattus no
22	48.4	2.8	1340	6	E13530	E13530 3'untransla
23	48.4	2.8	1844	6	E16086	E16086 cDNA encodi
24	48.4	2.8	2502	6	AR103240	AR103240 Sequence
25	48.4	2.8	2502	6	BD000103	BD000103 Different
26	48.4	2.8	2502	9	D82347	D82347 Homo sapien
27	48.4	2.8	6804	9	AB009997	AB009997 Homo sapi
28	48.4	2.8	7251	9	AB018693	AB018693 Homo sapi
29	48.4	2.8	152306	9	AC013733	AC013733 Homo sapi
30	48.4	2.8	165020	9	AC073977	AC073977 Homo sapi
31	47.4	2.8	151113	9	AC093858	AC093858 Homo sapi
32	47.4	2.8	188000	2	AC007895	AC007895 Homo sapi
33	46.6	2.7	179968	9	AC026228	AC026228 Homo sapi
34	46.4	2.7	1015	3	AF020280	AF020280 Dictyoste
35	45.6	2.7	527	3	DDITND312	K02644 Slime mold
36	45.6	2.7	1369	10	MMTPILR	X53333 Mouse mRNA
37	45.6	2.7	160759	2	AC117082	AC117082 Dictyoste
38	45.4	2.7	33404	2	AC115684	AC115684 Dictyoste
39	45.2	2.6	1830	10	BC018241	BC018241 Mus muscu
40	45.2	2.6	185004	2	AC107844	AC107844 Mus muscu
41	45.2	2.6	195948	2	AC129326	AC129326 Mus muscu
42	44.8	2.6	1527	8	ATATN1	X92728 A.thaliana
43	44.8	2.6	293545	2	AC098276	AC098276 Rattus no
44	44.4	2.6	131665	2	CNS08CB9	AL845347 Oryza sat
45	44.4	2.6	173918	2	AC108353	AC108353 Rattus no

ALIGNMENTS

RESULT 1  
AF197954

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AF197954 2021 bp mRNA linear PRI 05-JAN-2001  
Homo sapiens adenocarcinoma antigen ART1 mRNA, complete cds.

AF197954.1 GI:11066257

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2021)

Nishizaka,S., Gomi,S., Harada,K., Oizumi,K., Itoh,K. and

Shichijo,S.

A new tumor-rejection antigen recognized by cytotoxic T lymphocytes

infiltrating into a lung adenocarcinoma  
Cancer Res. 60 (17), 4830-4837 (2000)  
20441578  
MEDLINE  
10987294  
PUBMED  
2 (bases 1 to 2021)  
REFERENCE  
Nishizaka,S., Gomi,S., Shichiho,S., Harada,K., Kawano,K. and  
Itoh,K.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (24-Oct-1999) Immunology, Kurume University School of  
JOURNAL  
Medicine, Asahi-Machi 67, Kurume, Fukuoka 830-0011, Japan  
FEATURES  
Location/Qualifiers  
source  
1..2021  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
363..1607  
/note="T cell recognized antigen"  
/codon\_start=1  
/product="adenocarcinoma antigen Art1"  
/protein\_id="AAG28523.1"  
/db\_xref="gi:11066258"  
translation="MNLYRYWGEIPISSQTNRRSLDPREFRLVEVHPPLHOPSA  
NKPKPTMLDIPSPESLT IHT IOLIOHNRRLRLIATAQAOQOQTGKVKTESSEPL  
PSCGSPPLPDDLLPLDCKKNAPFQIRHSDPESDFYRGKPEVTELSWHSCRQLLYQ  
AVATLHAGDFDCANESVLETLTDVAHEYCLKFTKLRFVDFRARGUGQTFPPDMEQ  
VFHEVIGSVLSLQKFWQRIKDYHMLQISKOLSEYERIVRPERATEDAKPVKIK  
EPPVSDITFPVSELEADLSGDSQSLPMGLGAQSERPPSNLEVEASPAQSSAEVNAS  
PLMNHVKMPEQSEEGNVSGHVLGSDVFEPMMSGMSEAGIPQSPDDSDSYGSHS  
TDSLNGSSPVFNQCKKRMKI"  
BASE COUNT 510 a 523 c 489 g 499 t  
ORIGIN  
Query Match 98.9%; Score 1691.8; DB 9; Length 2021;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1693; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ACGCGATCCTTGCCCTCAGGCTCTCGAGGTCCAGACGCCGCCCGCTCTGCGGACG 60  
DB 23 ACGCGATCCTTGCCCTCAGGCTCTCGAGGTCCAGACGCCGCCCGCTCTGCGGACG 82  
QY 61 CAGCAGTGAATAGTGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 83 CAGCAGTGAATAGTGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 142  
QY 121 GCTGCCCTGACGCTCAGGCGACCTCAGACGCCCTGCTGATGGCGCTCGCGCGCGGACG 180  
DB 143 GCTGCCCTGACGCTCAGGCGACCTCAGACGCCCTGCTGATGGCGCTCGCGCGCGGACG 202  
QY 181 GTGACCGAGGAACCCCTGGAGGACCTGGGCGATTCCTTGGGCTCCGCTGCTCTTCG 240  
DB 203 GTGACCGAGGAACCCCTGGAGGACCTGGGCGATTCCTTGGGCTCCGCTGCTCTTCG 262  
QY 241 TGCTCTTTGCGGCAAGGATCTCACATTCAGTCTTTGAGCCACACAGAAATGCTGGCA 300  
DB 263 TGCTCTTTGCGGCAAGGATCTCACATTCAGTCTTTGAGCCACACAGAAATGCTGGCA 322  
QY 301 TTTGATAAATGTTTGTGAAGTGAAGACATATGGACAATGAATCTGCAAGATACTG 360  
DB 323 TTTGATAAATGTTTGTGAAGTGAAGACATATGGACAATGAATCTGCAAGATACTG 382  
QY 361 GGGAGAGATACCAATATATCAAGCCACACACAGAAAGTTCCCTTCGATTTGCTCCACG 420  
DB 383 GGGAGAGATACCAATATATCAAGCCACACACAGAAAGTTCCCTTCGATTTGCTCCACG 442  
QY 421 GGAGTTCGCTGTTGGAGTCCATGACCCCTGCACCAACCCCTCAGCCCAACAGCC 480  
DB 443 GGAGTTCGCTGTTGGAGTCCATGACCCCTGCACCAACCCCTCAGCCCAACAGCC 502  
QY 481 GAAGCCCCCCTACTGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGAT 540  
DB 503 GAAGCCCCCCTACTGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGAT 562  
QY 541 TCAGTTGATTGAGCACACCGACGCTTCGCAACCTTATTCGCACAGCTCAGGCCCGAGAA 600  
|||||

563 TCAGTTGATTTCAGCACAAACCGACGCTTTTCGCAACCTTATTCGCACAGCTCAGGCCCGAGAA 622  
QY 601 TCAGCAGCAGACAGAGAGTGTAAACACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTCGG 660  
|||||  
623 TCAGCAGCAGACAGAGAGTGTAAACACTGAAGAGAGTGAACCTCTTCCCTCGTCCCTCGG 682  
QY 661 GTCACCTCCTCTCCCTGATGACCTCCTGCTGCTTTAGATTGTAAAGATCCCAATGACACCAT 720  
|||||  
683 GTCACCTCCTCTCCCTGATGACCTCCTGCTGCTTTAGATTGTAAAGATCCCAATGACACCAT 742  
QY 721 CCAGATCCGGCAGCAGTGAACCCAGAGAGTGAACCTTTATCTGGGAAAGGGGAGGACCTGTGAC 780  
|||||  
743 CCAGATCCGGCAGCAGTGAACCCAGAGAGTGAACCTTTATCTGGGAAAGGGGAGGACCTGTGAC 802  
QY 781 TGAACCTGAGCTGGCAGCTCCTGTCGCGACGCTCTCTACAGCAGTGGCCACAACTCCTGCGC 840  
|||||  
803 TGAACCTGAGCTGGCAGCTCCTGTCGCGACGCTCTCTACAGCAGTGGCCACAACTCCTGCGC 862  
QY 841 CCAGCGGGCTTTGACTGTGCTGAATGAGAGTGTCTCTGGAGACCCCTAACTGATGTGGCACA 900  
|||||  
863 CCAGCGGGCTTTGACTGTGCTGAATGAGAGTGTCTCTGGAGACCCCTAACTGATGTGGCACA 922  
QY 901 TGAGTATTGCTTAAAGTTTACCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
|||||  
923 TGAGTATTGCTTAAAGTTTACCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 982  
QY 961 GGGACAGACTCCTTTTCTGATGTGATGAGCAGGATTTCCATGAAGTGGGTATTGGCAG 1020  
|||||  
983 GGGACAGACTCCTTTTCTGATGTGATGAGCAGGATTTCCATGAAGTGGGTATTGGCAG 1042  
QY 1021 TGTGCTCTCCTCCAGAGTTTCTGCGACGACCCGATCAAGGACTATCACAGTTACATGCT 1080  
|||||  
1043 TGTGCTCTCCTCCAGAGTTTCTGCGACGACCCGATCAAGGACTATCACAGTTACATGCT 1102  
QY 1081 ACAGATTAGTAAAGCACTCTCTGAAGATATGAAGGATTTGCTCAATCTCGAAGGCCAC 1140  
|||||  
1103 ACAGATTAGTAAAGCACTCTCTGAAGATATGAAGGATTTGCTCAATCTCGAAGGCCAC 1162  
QY 1141 AGAGCAGCTAAACCTGGAAGATCAAGGAGGAACTCTGAGGACGACATCACTTTTCCTGT 1200  
|||||  
1163 AGAGCAGCTAAACCTGGAAGATCAAGGAGGAACTCTGAGGACGACATCACTTTTCCTGT 1222  
QY 1201 CAGTGAAGAGTGGAGGCTGACCTTGTCTTGGAGACGACGACTGCTGCTTATGGAGTGTCT 1260  
|||||  
1223 CAGTGAAGAGTGGAGGCTGACCTTGTCTTGGAGACGACGACTGCTGCTTATGGAGTGTCT 1282  
QY 1261 TGGGGCTCAGAGCGAAGCTTCCCATCTAACTGGAGTTGAAGCTTCAACACGACGCTTC 1320  
|||||  
1283 TGGGGCTCAGAGCGAAGCTTCCCATCTAACTGGAGTTGAAGCTTCAACACGACGCTTC 1342  
QY 1321 AAGTGCAGAGTAAATGCTTCTCTCTTGGAACTGCGCCCATGTGAAATGGAGCTTCA 1380  
|||||  
1343 AAGTGCAGAGTAAATGCTTCTCTCTTGGAACTGCGCCCATGTGAAATGGAGCTTCA 1402  
QY 1381 AGAAGTGAAGAGCAATGCTCTGGGCAATGCTGCTGGGAGTGTGCTGGGAGTGTGCTTGGAGGA 1440  
|||||  
1403 AGAAGTGAAGAGCAATGCTCTGGGCAATGCTGCTGGGAGTGTGCTGGGAGTGTGCTTGGAGGA 1462  
QY 1441 GCCTATGTCAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAG 1500  
|||||  
1463 GCCTATGTCAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAG 1522  
QY 1501 CTATGGTTTCCCACTCCACTGACGCTCATGGGCTCTCTCCCTGTCTTTCACACCGCGCTG 1560  
|||||  
1523 CTATGGTTTCCCACTCCACTGACGCTCATGGGCTCTCTCCCTGTCTTTCACACCGCGCTG 1582  
QY 1561 CAAGAAGAGTGAAGAAATATATAAGAAAGAGGAGATGTTTGTTCAGACCTACT 1620  
|||||  
1583 CAAGAAGAGTGAAGAAATATATAAGAAAGAGGAGATGTTTGTTCAGACCTACT 1642  
QY 1621 AGACCCCAACAGAAAGGTTTGTGATTAGATCTGTTTCCCTTAAAAATTTGATTGACATCC 1680  
|||||  
1643 AGACCCCAACAGAAAGGTTTGTGATTAGATCTGTTTCCCTTAAAAATTTGATTGACATCC 1702



```

Qy 1681 TGTTCTTAAAAAAA 1695
Db 1703 TGTCTTAAACACAA 1717

RESULT 2
AB018307
LOCUS
DEFINITION Homo sapiens mRNA for KIAA0764 protein, complete cds.
ACCESSION AB018307
VERSION AB018307.1 GI:3882248
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N.,
Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 5 (5), 277-286 (1998)
MEDLINE 99087487
REFERENCE 2 (bases 1 to 4261)
AUTHORS Ohara, O., Suyama, M., Nagase, T., Ishikawa, K. and Kikuno, R.
DIRECT SUBMISSION
TITLE Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp. Tel: +81-438-52-3913,
Fax: +81-438-52-3914)
FEATURES
Location/Qualifiers
1..4261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hk04750"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/dev_stage="adult"
1..4261
/gene="KIAA0764"
344..1588
/gene="KIAA0764"
/codon_start=1
/product="KIAA0764 protein"
/protein_id="BAA34484.1"
/db_xref="GI:3882248"
/translation="MNLQRYWGEIPISSOTNRSLFLLPREFLRVEVHPPLHQPSPA
NKPPPTMLDIPSECSLTHTIQLIQRRLRLNLTAAQNOQOQTEGVKTESEPL
PSCPSPPLDLLPLDCKNPAPFIRHSDPESDFYRGKGEPTVLSLHWSRQLLYQ
AVATILAHGDFCANESVLETLTVAHEYCLKFTKLRLFAVDREARLGTPTPPVMEQ
VFEHVGTSVLSCFKWHRIDYHMYLQISKLSSEYERIVNPEKATEDAKPVKIK
EFVSDITPSEELRADLSGDSQSLPMGLAQSERFPNLEVEASPOASAEVNAS
PLWNLAHVMEQSEEGNVSGHVLGSDVFPEPMSGMEAGIPQSPDDSDSYGSHS
TDSLMSGSPVFVNRCKMRKI"
BASE COUNT 1171 a 966 c 897 g 1227 t
ORIGIN
Query Match 98.8%; Score 1690.2; DB 9; Length 4261;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1692; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AC CGCATCTTGCCTCAGGCCTCTCGAGTCCAGACAGCGCGCCAGCCGCGCTCTGGAGC 60
Db 1 AC CGCATCTTGCCTCAGGCCTCTCGAGTCCAGACAGCGCGCCAGCCGCGCTCTGGAGC 63

Qy 61 CAGCAGTGAATAGTGGTACCTCTGTCGTTGCTCGGTTGAGTCCAGACCTCCCGCTTCC 120
Db 64 CAGCAGTGAATAGTGGTACCTCTGTCGTTGCTCGGTTGAGTCCAGACCTCCCGCTTCC 123

```

```
|||||
Db 1204 CAGTCAGAGAGCTGGAGGCTGACCTTGTCTGGAGACAGTCACTGCTATGGAGTCT 1263
Qy 1261 TGGGCTCAGAGCAAGCTTCCATCTAAGCTGAGGTTGAAGCTTCCACACAGGCTTC 1320
Db 1264 TGGGCTCAGAGCAAGCTTCCATCTAAGCTGAGGTTGAAGCTTCCACACAGGCTTC 1323
Qy 1321 AAGTCAGAGTAATGCTTCTCTCTTTGGAATCTGGCCATGTGAAATGGAGCTCA 1380
Db 1324 AAGTCAGAGTAATGCTTCTCTCTTTGGAATCTGGCCATGTGAAATGGAGCTCA 1383
Qy 1381 AGAAGTGAAGCAAGCAATGCTCTGGCATGGTGTCTGGGAGTGAATGCTTCGAGGA 1440
Db 1384 AGAAGTGAAGCAAGCAATGCTCTGGCATGGTGTCTGGGAGTGAATGCTTCGAGGA 1443
Qy 1441 GCCTATGTCAGGATAGTGAAGCTGCTCCTCAGAGGCTCATGACTCAGATAGCAG 1500
Db 1444 GCCTATGTCAGGATAGTGAAGCTGCTCCTCAGAGGCTCATGACTCAGATAGCAG 1503
Qy 1501 CTATGGTTCACCTCCACTGACGCTCATGGGCTCTCCCTCTTTTCAACACGCGTG 1560
Db 1504 CTATGGTTCACCTCCACTGACGCTCATGGGCTCTCCCTCTTTTCAACACGCGTG 1563
Qy 1561 CAAGAAGAGTGAAGCAATATATAAGGAAAGAGGAGATGTTTGTCTCAGACCTACT 1620
Db 1564 CAAGAAGAGTGAAGCAATATATAAGGAAAGAGGAGATGTTTGTCTCAGACCTACT 1623
Qy 1621 AGACCCACAGAAAGGTTTGTATTAGATCTGTTTCTTAAATTTGATTGACTCC 1680
Db 1624 AGACCCACAGAAAGGTTTGTATTAGATCTGTTTCTTAAATTTGATTGACTCC 1683
Qy 1681 TGTCTTAAATAAAAA 1695
Db 1684 TGTCTTAAACACAA 1698

RESULT 3
AK000329 2045 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens cDNA FLJ20322 fis, clone HEP09455, highly similar to
DEFINITION AB018307 Homo sapiens mRNA for KIAA0764 protein.
ACCESSION AK000329
VERSION AK000329.1 GI:7020342
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
clone:HEP09455.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
TITLE NEDO human cDNA sequencing project
REFERENCE 2 (bases 1 to 2045)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
JOURNAL University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source 1..2045
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP09455"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
1..2045
misc_feature
/note="highly similar to AB018307 Homo sapiens mRNA for
KIAA0764 protein"
BASE COUNT 529 a 526 c 489 g 501 t
ORIGIN
Query Match 98.7%; Score 1688.6; DB 9; Length 2045;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1691; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ACAGGATCTTCCTCAGGCTCTCAGGCTCTCAGAGTCCAGACAGCCGCCAGCCGCTCTGCGACG 60
Db 32 ACAGGATCTTCCTCAGGCTCTCAGGCTCTCAGAGTCCAGACAGCCGCCAGCCGCTCTGCGACG 91
Qy 61 CAGCAGTGAATAGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 92 CAGCAGTGAATAGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 151
Qy 121 GGCTGCCCTGAACGCTCAGGCGACCTCAGGACCTCTGTGTTGGCGCCCTGCGCGCGGACG 180
Db 152 GGCTGCCCTGAACGCTCAGGCGACCTCAGGACCTCTGTGTTGGCGCCCTGCGCGCGGACG 211
Qy 181 GTGACGAGGAAACCCCTGGAGGACTTGGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 212 GTGACGAGGAAACCCCTGGAGGACTTGGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 271
Qy 241 TGCTCTCTCGGCAAGGATCTCAGTATATCAGTCTTTGACGACAGATGCTGCGCA 300
Db 272 TGCTCTCTCGGCAAGGATCTCAGTATATCAGTCTTTGACGACAGATGCTGCGCA 331
Qy 301 TTTGATAAATGTTTGTGAAGTCTGAAGACATATGACATGAATCTGCAAGATACG 360
Db 332 TTTGATAAATGTTTGTGAAGTCTGAAGACATATGACATGAATCTGCAAGATACG 391
Qy 361 GGGAGATACCAATATATCAAGCCAGACCAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db 392 GGGAGATACCAATATATCAAGCCAGACCAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 451
Qy 421 GGAGTTCCTGCTGTGGAAGTCCATGACCAACCCCTGCACCAACCCCTCAGCAACAGCC 480
Db 452 GGAGTTCCTGCTGTGGAAGTCCATGACCAACCCCTGCACCAACCCCTCAGCAACAGCC 511
Qy 481 GAAGCCCCCCTACTATGCTGGACATCCCTCTCAGAGCCATGTAGTCTACCATCCATACGAT 540
Db 512 GAAGCCCCCCTACTATGCTGGACATCCCTCTCAGAGCCATGTAGTCTACCATCCATACGAT 571
Qy 541 TCAGTTGATTCAGCAACAGCGCTCTCGCAACCTTATGCGCACAGCTTACGAGCCAGCA 600
Db 572 TCAGTTGATTCAGCAACAGCGCTCTCGCAACCTTATGCGCACAGCTTACGAGCCAGCA 631
Qy 601 TCAGCAGCAGCAGAGAGTGTAAACCTGAAGAGAGTGAACCTCTCTCTCTCTCTCTCTCTCT 660
Db 632 TCAGCAGCAGCAGAGAGTGTAAACCTGAAGAGAGTGAACCTCTCTCTCTCTCTCTCTCTCT 691
Qy 661 GTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 692 GTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 751
Qy 721 CCAGATCCGCGACAGTGAACCCAGAGAGTACTTTTATCGTGGGAAAGGGAACCTGTGAC 780
Db 752 CCAGATCCGCGACAGTGAACCCAGAGAGTACTTTTATCGTGGGAAAGGGAACCTGTGAC 811
Qy 781 TGAAGTCACTGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
Db 812 TGAAGTCACTGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 871
```



QY 671 CTCCTGATGACCTCCTGCTTTAGATTGTAAGAAATCCCAATGACCACTTCAGATCCGG 730  
Db 488 CTCCTGATGACCTCCTGCTTTAGATTGTAAGAAATCCCAATGACCACTTCAGATCCGG 547  
QY 731 CACAGTGACCCAGAGAGTACTTTTATCGTGGGAAGGGGAACTGCTGACTCAACTCAGC 790  
Db 548 CACAGTGACCCAGAGAGTACTTTTATCGTGGGAAGGGGAACTGCTGACTCAACTCAGC 607  
QY 791 TGGCACTCCTGCTGGCAGCTCCTTACCAGGAGTGGCCACAAATCCTGCCCCACCGGGC 850  
Db 608 TGGCACTCCTGCTGGCAGCTCCTTACCAGGAGTGGCCACAAATCCTGCCCCACCGGGC 667  
QY 851 TTTGACTGCTTAATGAGAGTCTCTGGAGACCCCTAACTGATGTCGACATGAGTATTC 910  
Db 668 TTTGACTGCTTAATGAGAGTCTCTGGAGACCCCTAACTGATGTCGACATGAGTATTC 727  
QY 911 CTTAAGTTTACCAAGTCTGCTGGTTTCTGCTGGACCGGGAGCGCCGCTGGACAGACT 970  
Db 728 CTTAAGTTTACCAAGTCTGCTGGTTTCTGCTGGACCGGGAGCGCCGCTGGACAGACT 787  
QY 971 CTTTTCTGATGATGGAGCAGGTATTCCATGAAGTGGGTATTGGCAGTGTGCTCTCC 1030  
Db 788 CTTTTCTGATGATGGAGCAGGTATTCCATGAAGTGGGTATTGGCAGTGTGCTCTCC 847  
QY 1031 CTCAGAGATTCTGGCAGACCGCATCAAGGACTATCACAGTTACATGCTACAGATTAGT 1090  
Db 848 CTCCTAAAGTTCTGGCAGACCGCATCAAGGACTATCACAGTTACATGCTACAGATTAGT 907  
QY 1091 AAGCAACTCTCGAAGATATGAAGGATTTGCTGCTGGACCGGGAGCGCCGCTGGACAGACT 1150  
Db 908 AAGCAACTCTCGAAGATATGAAGGATTTGCTGCTGGACCGGGAGCGCCGCTGGACAGACT 967  
QY 1151 AAACCTGTGAAGATCAAGGAGGAACTGTGAGCAGATCACTTTCTGCTAGTGAGGAG 1210  
Db 968 AAACCTGTGAAGATCAAGGAGGAACTGTGAGCAGATCACTTTCTGCTAGTGAGGAG 1027  
QY 1211 CTGGAGGTGACCTTCTCTGGAGACCGCATCACTGCTGCTATGGGAGTGTGGGGCTCAG 1270  
Db 1028 CTGGAGGTGACCTTCTCTGGAGACCGCATCACTGCTGCTATGGGAGTGTGGGGCTCAG 1087  
QY 1271 AGCGAAGCTTCCCATCTAACTGAGGTTGAAGCTTCACACAGGCTTCAAGTGCAGAG 1330  
Db 1088 AGCGAAGCTTCCCATCTAACTGAGGTTGAAGCTTCACACAGGCTTCAAGTGCAGAG 1147  
QY 1331 GFAAATGCTTCTCTCTTTGGAATCTGGCCCATGTGAAAATGAGCCCTCAAGAAAGTAA 1390  
Db 1148 GFAAATGCTTCTCTCTTTGGAATCTGGCCCATGTGAAAATGAGCCCTCAAGAAAGTAA 1207  
QY 1391 GAAGCAATGCTCTGGGCAATGCTGCTGGGCACTGATGCTTCGAGGAGCCTATGTCA 1450  
Db 1208 GAAGCAATGCTCTGGGCAATGCTGCTGGGCACTGATGCTTCGAGGAGCCTATGTCA 1267  
QY 1451 GGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCC 1510  
Db 1268 GGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCC 1327  
QY 1511 CACTCCACTGACAGCCTCATGGGGTCTCCCTGTTTCAACAGCGCTGCAAGAAAGAGG 1570  
Db 1328 CACTCCACTGACAGCCTCATGGGGTCTCCCTGTTTCAACAGCGCTGCAAGAAAGAGG 1387  
QY 1571 ATGAGGAAATATATAAGGAAAGAGGAGATGTTTTGTCAGACCTACTAGACCCACA 1630  
Db 1388 ATGAGGAAATATATAAGGAAAGAGGAGATGTTTTGTCAGACCTACTAGACCCACA 1447  
QY 1631 GAAAAGGTTTTGTTATAGAA 1651  
Db 1448 GAAAAGGTTTTGTTATATAA 1468

RESULT 5  
AF224759  
LOCUS  
DEFINITION Homo sapiens adenocarcinoma antigen ART1/P17 mRNA, complete cds.  
2869 bp mRNA linear PRI 07-JAN-2001

AF224759  
AF224759.1 GI:12043739  
KEYWORDS  
SOURCE Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 2869)  
Nishizaka, S., Gomi, S., Harada, K., Oizumi, K., Itoh, K. and  
Shichijo, S.  
TITLE  
A new tumor-rejection antigen recognized by cytotoxic T lymphocytes  
infiltrating into a lung adenocarcinoma  
Cancer Res. 60 (17), 4830-4837 (2000)  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
2 (bases 1 to 2869)  
Nishizaka, S., Gomi, S., Harada, K., Oizumi, K., Itoh, K. and  
Shichijo, S.  
TITLE  
Direct Submission  
JOURNAL  
SUBMITTED (14-JAN-2000) Immunology, Kurume University School of  
Medicine, Asahi-machi 67, Kurume, Fukuoka 830-0011, Japan  
FEATURES  
Location/Qualifiers  
Source  
1..2869  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="adenocarcinoma"  
1198..2436  
/codon\_start=1  
/product="antigen ART1/p17"  
/protein\_id="AAG47636.1"  
/db\_xref="GI:12043740"  
/translation="MLRYWGEIPISSTNRSFDLLPREFRLVEVHPDPLHPSANK  
PKPTMLDIPSEPCSLTHTTQLIHNRLNLATAQONQQTGKTESESEPLPS  
CPSPPLPDLPLDCKNPAPFOIRHSDPSDFVKGPEVTELSWISCRLLYQAV  
ATILAHAGFCANESVLETLDVAHEVCLKTLRLFAVDREARLGGTGPFDVMEQVF  
HEVIGSVLSLQKFWHRIKDYHSLMQLSKLSEERYINPEKATADAPVKKEE  
PVSDDTFVSEELDLASGQSLPMGVIGLSDVFEFPMSEAGIPQSPDDSDSYSGHST  
NLARVMEPEQSEEGNVSGHVLGSDVFEFPMSEAGIPQSPDDSDSYSGHST  
SLMGSPVFNQRCRKRKI"  
BASE COUNT 769 a 673 c 655 g 772 t  
ORIGIN  
Query Match 78.4%; Score 1340.8; DB 9; Length 2869;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 344 AATCTGCAAGATACTGGGAGAGATACCAATATCATCATCAAGCCAGACCAAGAGATTCC 403  
Db 1195 AAAATGTTGATATCTGGGAGAGATACCAATATCATCATCAAGCCAGACCAAGAGATTCC 1254  
QY 404 TTCGATTTGCTCCACGGGAGTTCGCTCTGGTGAAGTCCATGACCCACCCCTGCACCAA 463  
Db 1255 TTCGATTTGCTCCACGGGAGTTCGCTCTGGTGAAGTCCATGACCCACCCCTGCACCAA 1314  
QY 464 CCCTCAGCCCAACAGCCGAAAGCCGCCCTACTATGCTGGACATCCCTCAGAGCCATGATG 523  
Db 1315 CCCTCAGCCCAACAGCCGAAAGCCGCCCTACTATGCTGGACATCCCTCAGAGCCATGATG 1374  
QY 524 CTCACCATCCATGATTCAGTTTCTGATTCAGCAGACACCGACCTCTTCGCAACTTATTC 583  
Db 1375 CTCACCATCCATGATTCAGTTTCTGATTCAGCAGACACCGACCTCTTCGCAACTTATTC 1434  
QY 584 ACAGCTCAGGCGCCAGAAATCAGCAGCAGACAGAGGTGTAAAACCTGAAGAGAGTGAACCT 643  
Db 1435 ACAGCTCAGGCGCCAGAAATCAGCAGCAGACAGAGGTGTAAAACCTGAAGAGAGTGAACCT 1494  
QY 644 CTTCCCTCGTCCCTGGGTGCTACCTCTCCTGATGACCTCCCTGCTTGTAGATTGTAAG 703  
Db 1495 CTTCCCTCGTCCCTGGGTGCTACCTCTCCTGATGACCTCCCTGCTTGTAGATTGTAAG 1554  
QY 704 AATCCCAATGACCATTCAGATCCGGCAGCAGTGCACCGACAGAGTGACCTTTTATCGTGG 763  
Db 1555 AATCCCAATGACCATTCAGATCCGGCAGCAGTGCACCGACAGAGTGACCTTTTATCGTGG 1614

QY	764	AAAGGGGAACCTGTGACTGAACCTCAGCTGGCACTCCTGTGCGCAGCTCCTCTACCAAGCA	823
Db	1615	AAAGGGGAACCTGTGACTGAACCTCAGCTGGCACTCCTGTGCGCAGCTCCTCTACCAAGCA	1674
QY	824	GTGGCCCAATCCTGGCCACGCGGCTTTGACTGTGCTATGAGAGTGTCTCTGGAGACC	883
Db	1675	GTGGCCCAATCCTGGCCACGCGGCTTTGACTGTGCTATGAGAGTGTCTCTGGAGACC	1734
QY	884	CTAACTGATGTGGCAGATGATTGCTTAAAGTTTACCAAGTTGCTCGCTTTTCTGTG	943
Db	1735	CTAACTGATGTGGCAGATGATTGCTTAAAGTTTACCAAGTTGCTCGCTTTTCTGTG	1794
QY	944	GACCGGAGGCGCGCTGGGACAGACCTCTTTTCTGTGATGTGATGGAGCAGGTATTCAT	1003
Db	1795	GACCGGAGGCGCGCTGGGACAGACCTCTTTTCTGTGATGTGATGGAGCAGGTATTCAT	1854
QY	1004	GAGTGGGTATGTCAGTGTCTCTCCCTCAGAAAGTTCTGGCAGCAGCCGATCAAGGAC	1063
Db	1855	GAGTGGGTATGTCAGTGTCTCTCCCTCAGAAAGTTCTGGCAGCAGCCGATCAAGGAC	1914
QY	1064	TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGATATGAAGGATTGTC	1123
Db	1915	TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGATATGAAGGATTGTC	1974
QY	1124	AATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGC	1183
Db	1975	AATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGC	2034
QY	1184	GACATCACTTTTCTGCTGAGTGGAGCTGGAGCTGACCTTCTCTTTTGGAGACCACTCA	1243
Db	2035	GACATCACTTTTCTGCTGAGTGGAGCTGGAGCTGACCTTCTCTTTTGGAGACCACTCA	2094
QY	1244	CTGCCTATGGAGTGTCTGGGCTCAGAGCGAAGCTTCCCATCTAACTCGAGGAGTTGAA	1303
Db	2095	CTGCCTATGGAGTGTCTGGGCTCAGAGCGAAGCTTCCCATCTAACTCGAGGAGTTGAA	2154
QY	1304	GCTTACACAGGCTTCAAGTGCAGAGGTAAGTCTTCTCTTTTGGAACTCTGGCCCAT	1363
Db	2155	GCTTACACAGGCTTCAAGTGCAGAGGTAAGTCTTCTCTTTTGGAACTCTGGCCCAT	2214
QY	1364	GTGAAATGGAGCTCAAGAAAGTGAAGAGGCAATGTCTCTGGGCAATGTCTGGGCT	1423
Db	2215	GTGAAATGGAGCTCAAGAAAGTGAAGAGGCAATGTCTCTGGGCAATGTCTGGGCT	2274
QY	1424	AGTGATGCTCTCGAGAGGCTTATGTGAGGATGAGTGAAGTGGGATTCCTCAGAGCCT	1483
Db	2275	AGTGATGCTCTCGAGAGGCTTATGTGAGGATGAGTGAAGTGGGATTCCTCAGAGCCT	2334
QY	1484	GATGACTCAGATAGCAGTATGGTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT	1543
Db	2335	GATGACTCAGATAGCAGTATGGTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT	2394
QY	1544	GTTTTCAACGAGCGCTCAGAGAGGATGAGGAAATATAAAGGAAAGAGGAGGATG	1603
Db	2395	GTTTTCAACGAGCGCTCAGAGAGGATGAGGAAATATAAAGGAAAGAGGAGGATG	2454
QY	1604	TTTTTCCAGACCTACTAGACCCCAACAGAAAGGTTTTTGTATTAGATCTGTTTCTCTTA	1663
Db	2455	TTTTTCCAGACCTACTAGACCCCAACAGAAAGGTTTTTGTATTAGATCTGTTTCTCTTA	2514
QY	1664	AAATGATTGATCTGCTGTTCTTAAAAAAA	1695
Db	2515	AAATGATTGATCTGCTGTTCTTAAAAACAA	2546

RESULT 6  
BC031447  
LOCUS  
DEFINITION Mus musculus, RIKEN cDNA 2610524B01 gene, clone MGC:25489  
IMAGE:4502314, mRNA, complete cds.  
ACCESSION BC031447  
VERSION BC031447.1 GI:21594424

KEYWORDS	MGC.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2616)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgaps-re@mail.nih.gov">cgaps-re@mail.nih.gov</a> Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a> Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a> Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 31 Row: k Column: 15 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction, Similarity but not identity to protein. Location/Qualifiers 1..2616 /organism="Mus musculus" /db_xref="taxon:10090" /clone="MGC:25489 IMAGE:4502314" /tissue_type="Eye, retina, mouse strain C57Bl/6" /clone_lib="NIH_MGC_94" /lab_host="DH10B" /note="vector: pCMV-Sport6" 267..1505 /codon_start=1 /product="RIKEN cDNA 2610524B01 gene" /protein_id="AAH31447.1" /db_xref="GI:21594425" /db_xref="LocusID:72195" /translation="MLRYWGEIPIPSGQTNRSSFDLLPRFRLVEVHDPPLHPSANK PKPTMLDIPSEPCSLTHTLIQHNRLSLIATATQSQOQTEGVKAESEPLPS CPGSPPLDPLDCKPNAPFOIRHSDPESDFYRGKGPVTELSWHSCROLLYQAV ATLAWHFGSCANESVLETLIDVAHEYCLKFTKLRFAYDREALLGTPPDVMEQVF HEVGISVLSQNFQWHRINDYTMQLQISQJSEERYIVNPEKATEDTPVKIRKEE VNSDITPPVSEELADIASGDSLPIGLVQAQSERFSPNLEVASPOAFAEYNASPL PWLAHYKMEPESEBEGNVSAHGVLGDFEFPMSGMEAGLPQSPDDSDSYGSHSTD SLMGSSPVFNQRCRKRMT"
CDS	BASE COUNT 651 a 682 c 620 g 663 t ORIGIN
Query Match	61.8%; Score 1057; DB 10; Length 2616;
Best Local Similarity	90.4%; Pred. No. 1.8e-291;
Matches 1129; Conservative	0; Mismatches 120; Indels 0; Gaps 0;
QY	344 AATCTGCAAGATACCTGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTC 403
Db	264 AAAATGCTGAGGTACTGGGAGAGATACCAATATCGGTCAAGCAGACACAGAGTTC 323
QY	404 TTGATCTGCTCCACCGGAGTTCGGTCTGGTGAAGTCCATGACCCACCCCTGCACAA 463
Db	324 TTTGATCTGCTCCACCGGAGTTCGGTCTGGTGAAGTCCATGACCCGCCCTGCACAG 383
QY	464 CCCTCAGCCAAACAGCCGAGCCGCCCACTATGCTGGACATCCCTCAGAGCCATGTAGT 523

Db 384 CCTCAGCCAAAGCCGAGCCGACGATGCTGGACATTCCTCAGAACCGGTGATG 443  
QY 524 CTCACATCATACGATTCAGTTCAGACACACCGACGCTTCGCAACCTTATTGCC 583  
Db 444 CTCACATCATACGATTCAGTTCAGACACACCGACGCTTCGCAACCTTATTGCC 503  
QY 584 ACAGCTCAGGCCGAGCAATCAGCAGCAGACAGAGGTGTAAACACTGAAGAGAGTGAACCT 643  
Db 504 ACAGCTCAGGCCGAGCAATCAGCAGCAGACAGAGGTGTGAAGGCTGAAGAGAGTGAACCT 563  
QY 644 CTTCCCTCGTGGCTGAGTACCTCCTCTCCCTGTAGACCTCCTCCCTTTAGATGTAAAG 703  
Db 564 CTTCCCTCGTGGCTGAGTACCTCCTCTCCCTGTAGACCTCCTCCCTTTAGATGTAAAG 623  
QY 704 ATCCCAATGACCAATTCAGATCCGCGACAGTACCGACAGAGTACCTTTATCGTGGG 763  
Db 624 ATCCCAATGACCAATTCAGATCCGCGACAGTACCGACAGAGTACCTTTATCGTGGG 683  
QY 764 AAAGGGAGACCTGTGACTGAATCAGCTGGCACTCCTGTGGCAGCTCCTCTACCAAGCA 823  
Db 684 AAAGGAGACCTGTGACAGAGTGAAGTGGCACTCCTGCCGACGCTCCTATCAAGCA 743  
QY 824 GTGGCCACAATCCTGGCCACGCGGCTTTGACTGTGCTAATGAGAGTGTCTGGAGACC 883  
Db 744 GTGGCCACAATCCTGGCCACGAGGCTTTGAGTGTGCTAATGAAAGTGTCTGGAGACC 803  
QY 884 CTAACGTATGTGGACATAGTATTGCTTAAAGTTTACCAAGTGTCTGGTCTGCTGTG 943  
Db 804 CTAACGTATGTGGACATAGTATTGCTTAAAGTTTACCAAGTGTCTGGTCTGCTGTG 863  
QY 944 GACCGGAGGCGGCTGGGACAGACTCCTTTCTGTATGTGATGAGCAGGTATTCAT 1003  
Db 864 GACCGGAGGCGGCTGGGACAGACTCCTTTCTGTATGTGATGAGCAGGTATTCAT 923  
QY 1004 GAAAGTGGTATTTGGCAGTGTGCTCTCCCTCAGAAAGTCTTGGCAGCACCAGATCAAGAC 1063  
Db 924 GAAAGTGGCAGTGTGCTCTCCCTCAGAAAGTCTTGGCAGCACCAGATCAAGAC 983  
QY 1064 TATCAGTATACATCTACAGATAGTAAAGCACTCTCTGAAGATATGAAGGATTCG 1123  
Db 984 TATCAGTATACATCTACAGATAGTAAAGCACTCTCTGAAGATATGAAGGATTCG 1043  
QY 1124 AATCTGAGAGGCGCACAGAGACGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGC 1183  
Db 1044 AATCCCGAGAGGCGCACAGAGACACTAAACCTGTGAAGATCAAGGAGAACCTGTGAGC 1103  
QY 1184 GACATCACTTTCTGTGATGAGGAGTGGAGGCTGACCTTGTCTTGGAGACAGTCA 1243  
Db 1104 GACATCACTTTCTGTGATGAGGAGTGGAGGAGCTTGTCTTGGAGATCAGTCC 1163  
QY 1244 CTGCTATGGAGTGTCTGGGCTCAGAGCGAAGCTTCCCATCTAACCTGGAGTTCGA 1303  
Db 1164 TTACCCATTTGGGCTCTCGGGCTCAGAGTGAAGCTTCCCATCTAACCTGGAGTTCGA 1223  
QY 1304 GCTTCCACACAGGCTTCAAGTGCAGAGTAAATGCTTCTCTCTTTGGAATCTGGCCCAT 1363  
Db 1224 GCTTCCACACAGGCTTCAAGTGCAGAGTAAATGCTTCTCTCTTTGGAATCTGGCCCAT 1283  
QY 1364 GTGAANAATGGAGCTTCAAGAAAGTGAAGAGCAATGCTTGGGCAATGCTGGGCTGGG 1423  
Db 1284 GTGAANAATGGAGCTTCAAGAAAGTGAAGAGCAATGCTTGGGCAATGCTGGGCTGGG 1343  
QY 1424 AGTATGTCTTCGAGGAGCTTATGTACAGGATAGTGAAGCTGGGATTCCTCAGAGCCCT 1483  
Db 1344 AGTATGTCTTCGAGGAGCTTATGTACAGGATAGTGAAGCTGGGATTCCTCAGAGCCCT 1403  
QY 1484 GATGACTCAGATAGCAGTATGGTTCACCTCCACTGACAGCCCTCATGGGGTCTCCCT 1543  
Db 1404 GACGACTCAGAGCAGTATGGTTCACCTCCACTGACAGCCCTCATGGGGTCTCCCT 1463  
QY 1544 GTTTTCACCCGCTGCAAGAGGATGAGGAAATATATAAGGAAA 1592

Db 1464 GTGTTCAATCAGCGCTGCAGGAAGAGATGAGGAAGATTTAACTGCAGA 1512

RESULT 7  
AX013061  
LOCUS AX013061 1046 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from Patent WO9954461.  
ACCESSION AX013061  
VERSION AX013061.1 GI:10040227  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1046)  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pilarczyk,C.  
TITLE Human nucleic acid sequences of endometrium tumour tissue  
JOURNAL Patent: WO 9954461-A 1 28-OCT-1999;  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILARSKY CHRISTIAN (DE)  
FEATURES  
source Location/Qualifiers  
1..1046  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 240 a 304 c 262 g 240 t  
ORIGIN

Query Match 49.7%; Score 850.8; DB 6; Length 1046;  
Best Local Similarity 96.1%; Pred. No. 2.2e-232;  
Matches 949; Conservative 0; Mismatches 27; Indels 12; Gaps 7;

QY 1 ACGGGATCCTTGGCTCAGGCTCTCGAGGTCCAGACAGCCGCGCTCTCGGAGC 60  
Db 46 ACGGGATCCTTGGCTCAGGCTCTCGAGGTCCAGACAGCCGCGCTCTCGGAGC 105  
QY 61 CAGCAGTGAATAGTGTGTACCTCTCTCTCGGTTCAGGTCCAGACCTCCCGCTTTC 120  
Db 106 CAGCAGTGAATAGTGTGTACCTCTCTCTCGGTTCAGGTCCAGACCTCCCGCTTTC 165  
QY 121 GGCTGCCCTGAAGCTCAGGACCTCAGGACCTGTGATGGGCGCTCGCGGGGAGC 180  
Db 166 GGCTGCCCTGAAGCTCAGGACCTCAGGACCTGTGATGGGCGCTCGCGGGGAGC 225  
QY 181 GTGACCGAGGAACCCCTGGAGGACTTGGGCTTCCCTTGGGCTCCGCTGTCTTCG 240  
Db 226 GTGACCGAGGAACCCCTGGAGGACTTGGGCTTCCCTTGGGCTCCGCTGTCTTCG 285  
QY 241 TGCTCCTTTC-GGGCAGGATCTCACATTTATCAGTCTTTGACCGACACAGAAATGCTTGC 299  
Db 286 TGCTCCTTTCGGGCAAGGATCTCACATTTATCAGTCTTTGACCGACACAGAAATGCTTGC 345  
QY 300 ATTTGATAAATTTGTTGTAAGTTCAGAGACATATGAGCAATGAATTCGAAAGATACT 359  
Db 346 ATTTGATAAATTTGTTGTAAGTTCAGAGACATATGAGCAATGAATTCGAAAGATACT 405  
QY 360 GGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCTTCGATTTGCTCCAC 419  
Db 406 GGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCTTCGATTTGCTCCAC 465  
QY 420 GGGAGTTCCGCTCTGGTGAAGTCCATGACCCACCCCTGCACCAACCTTCAGCCCAAGC 479  
Db 466 GGGAGTTCCGCTCTGGTGAAGTCCATGACCCACCCCTGCACCAACCTTCAGCCCAAGC 525  
QY 480 CGAAGCCCCCAGTATGCTGGACATCCCTCAGAGCCATGTAGTCTACCATTCATAGA 539  
Db 526 CGAAGCCCCCAGTATGCTGGACATCCCTCAGAGCCATGTAGTCTACCATTCATAGA 585  
QY 540 TTCAGTTGATTCAGCAACCGAGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCACAGA 599  
Db 586 TTCAGTTGATTCAGCAACCGAGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCACAGA 645



```
QY 600 ATCAGCAGCAGACAGAGGTGTAAAACTGAAGAGAGTGAACCTCTTCCCTCGCGCCCTG 659
Db 646 ATCAGCAGCAGACAGAGGTGTAAAACTGAAGAGAGTGAACCTCTTCCCTCGCGCCCTG 705
QY 660 GGTCACTCTCTCCCTGATGACCTCTCGCTTTAGATTGTGAAGAAATCCCAATCACCAT 719
Db 706 GGTCACTCTCTCCCTGATGACCTCTCGCTTTAGATTGTGAAGAAATCCCAATCACCAT 765
QY 720 TCCAGATCCGCGACAGTGAACCCAGAGAGTGAACCTTTATCGTGGGAAAGGGAACTGTGA 779
Db 766 TCCAGATCCGCGACAGTGAACCCAGAGAGTGAACCTTTATCGTGGGAAAGGGAACTGTGA 825
QY 780 CTGAACCTCAGTGGCAGCTCTGTGGCAGCTCTCTACCG-AGGCAGTGGCCACAACTCTG 838
Db 826 CTGAACCTCAGTGGCAGCTCTGTGGCAGCTCTCTACCGAGGAGTGGCAGCAATCTG 885
QY 839 GCCCAGC-CGGGCTTTGACTGTGCTAATGAGAGTGTCTGG-AGACCCCTAACT--GATG 893
Db 886 GCCAACGGCGGCTTTGACTGTGCTAATGAGAGTGTCTGGGAAAGCCCTAACTGTATGT 945
QY 894 TGGCAGATGATGATTTG--CCTTAAGTTTACCAAGTGTCTGCG---TTTTGCTGTGGACCG 948
Db 946 TGGCAGATGATGATTTGCTTTAAGTTTACCAAGTGTCTGCGCTTTTGTCTGTGGCG 1005
QY 949 GGAGCGCGGCTGGGACAGACTCTCTTT 976
Db 1006 GGAAGCGCGGCTGGGAGAGACTCTCTTT 1033

RESULT 8
AC074091
DEFINITION Homo sapiens BAC clone RP11-158113 DNA linear PRI 10-JAN-2002
ACCESSION AC074091
VERSION AC074091.6 GI:17975414
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 204153)
AUTHORS Nguyen, C., Haglund, K., Dixon, R. and Doeber, A.
TITLE The sequence of Homo sapiens BAC clone RP11-158113
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 204153)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 204153)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 204153)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 204153)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

## COMMENT

On Dec 21, 2001 this sequence version replaced gi:16924152.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiensew@wustl.edu  
----- Summary Statistics  
-----  
Center project name: H\_NH0158113  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
All regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)  
VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-158113; actual end is at base position 204153 of RP11-158113.

A PCR only region exists between bases 64163 to 64357. A transposon exists in this clone that has been deleted from the finished sequence. This transposon would insert after base 38173.

## FEATURES

## source

1. 204153  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-158113"  
/clone\_lib="RPCI-11"

## misc\_feature

464  
/note="match to EST AI370811 (NID:g4149564) ta58c03.xl"

## misc\_feature

492..800  
/note="match to EST AA649214 (NID:g2575643) ns31f01.sl"

## misc\_feature

492..800  
/note="match to EST BF516222 (NID:g11601401)"

## misc\_feature

497..800  
/note="match to EST BI597621 (NID:g15490560)"

## misc\_feature

498..962  
/note="match to EST AI370811 (NID:g4149564) ta58c03.xl"

## misc\_feature

500..800  
/note="match to EST AV734630 (NID:g10852175)"

## misc\_feature

500..800  
/note="match to EST BE881448 (NID:g10330224)"

## misc\_feature

500..800  
/note="match to EST BG215442 (NID:g13741463)"

## misc\_feature

500..752  
/note="match to EST AA594514 (NID:g2409864) n194f03.sl"

## repeat\_region

500..520









**TITLE**

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 155250)

Worley,K.C.

Direct Submission

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GZSG

Center clone name: CH230-46303

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 88645 bases at least Q40

Consensus quality: 95723 bases at least Q30

Consensus quality: 100316 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1151: contig of 1151 bp in length  
\* 1152: gap of unknown length  
\* 1252: contig of 1315 bp in length  
\* 2567: gap of unknown length  
\* 2667: contig of 1557 bp in length  
\* 4224: gap of unknown length  
\* 4324: contig of 1082 bp in length  
\* 5405: gap of unknown length  
\* 5406: contig of 1064 bp in length  
\* 5506: gap of unknown length  
\* 6570: gap of unknown length  
\* 6670: contig of 1630 bp in length  
\* 8300: gap of unknown length  
\* 8400: contig of 1008 bp in length  
\* 9408: gap of unknown length  
\* 9508: contig of 1090 bp in length  
\* 10598: gap of unknown length  
\* 10698: contig of 1193 bp in length  
\* 11891: gap of unknown length  
\* 11991: contig of 1018 bp in length  
\* 13009: gap of unknown length  
\* 13109: contig of 1942 bp in length  
\* 15051: gap of unknown length  
\* 15151: contig of 1339 bp in length  
\* 16490: gap of unknown length  
\* 16590: contig of 1072 bp in length  
\* 17662: gap of unknown length  
\* 17762: contig of 1444 bp in length  
\* 19206: gap of unknown length  
\* 19306: contig of 1633 bp in length  
\* 20939: gap of unknown length  
\* 21039: contig of 1396 bp in length  
\* 22435: gap of unknown length  
\* 22535: contig of 1231 bp in length  
\* 23765: gap of unknown length  
\* 23865: contig of 1385 bp in length  
\* 25251: gap of unknown length  
\* 25351: contig of 1512 bp in length  
\* 26863: gap of unknown length  
\* 26963: contig of 1326 bp in length  
\* 28289: gap of unknown length

28389: contig of 1689 bp in length  
\* 30078: gap of unknown length  
\* 30178: contig of 1765 bp in length  
\* 31943: gap of unknown length  
\* 32043: contig of 1001 bp in length  
\* 33044: gap of unknown length  
\* 33144: contig of 1886 bp in length  
\* 35030: gap of unknown length  
\* 35130: contig of 1670 bp in length  
\* 36800: gap of unknown length  
\* 36900: contig of 2176 bp in length  
\* 39076: gap of unknown length  
\* 39176: contig of 1144 bp in length  
\* 40320: gap of unknown length  
\* 40420: contig of 1806 bp in length  
\* 42226: gap of unknown length  
\* 42326: contig of 1735 bp in length  
\* 44061: gap of unknown length  
\* 44161: contig of 2033 bp in length  
\* 46194: gap of unknown length  
\* 46294: contig of 1409 bp in length  
\* 47703: gap of unknown length  
\* 47803: contig of 1476 bp in length  
\* 49279: gap of unknown length  
\* 49379: contig of 1317 bp in length  
\* 50696: gap of unknown length  
\* 50796: contig of 1937 bp in length  
\* 52833: gap of unknown length  
\* 52833: contig of 2066 bp in length  
\* 54899: gap of unknown length  
\* 54999: contig of 3774 bp in length  
\* 58773: gap of unknown length  
\* 58873: contig of 2053 bp in length  
\* 60926: gap of unknown length  
\* 61026: contig of 1991 bp in length  
\* 63017: gap of unknown length  
\* 63117: contig of 3178 bp in length  
\* 66295: gap of unknown length  
\* 66395: contig of 3150 bp in length  
\* 69545: gap of unknown length  
\* 69645: contig of 2460 bp in length  
\* 72105: gap of unknown length  
\* 72205: contig of 2008 bp in length  
\* 74213: gap of unknown length  
\* 74313: contig of 1913 bp in length  
\* 76226: gap of unknown length  
\* 76326: contig of 2805 bp in length  
\* 79131: gap of unknown length  
\* 79231: contig of 2716 bp in length  
\* 81947: gap of unknown length  
\* 82047: contig of 4080 bp in length  
\* 86127: gap of unknown length  
\* 86227: contig of 5045 bp in length  
\* 91272: gap of unknown length  
\* 91372: contig of 2063 bp in length  
\* 93435: gap of unknown length  
\* 93535: contig of 6491 bp in length  
\* 100026: gap of unknown length  
\* 100126: contig of 3054 bp in length  
\* 103180: gap of unknown length  
\* 103280: contig of 3660 bp in length  
\* 106940: gap of unknown length  
\* 107040: contig of 3985 bp in length  
\* 111025: gap of unknown length  
\* 111125: contig of 6013 bp in length  
\* 117138: gap of unknown length  
\* 117238: contig of 4237 bp in length  
\* 121475: gap of unknown length  
\* 121575: contig of 4653 bp in length  
\* 126228: gap of unknown length  
\* 126328: contig of 5727 bp in length  
\* 132055: gap of unknown length  
\* 132155: contig of 6132 bp in length  
\* 138286: gap of unknown length

* 138287	138386:	gap of unknown length
Query Match	20.6%;	Score 352; DB 2; Length 155250;
Best Local Similarity	90.4%;	Pred. No. 6e-89;
Matches 376;	Conservative	0; Mismatches 40; Indels 0; Gaps 0;
QY	344	AATCTGCAAGACTACGGGAGAGATACCATTATCATCAACCCAGACAGAAAGTTCC 403
Db	81220	AAAATGCTGAGTACTGGGAGAGATCCCATTCATCAGCACAGACAAGAAGTTCC 81161
QY	404	TTCGATTCTCCCAACCGGAGTTCCGTCTGTTGAAGTCCATGATGCCACCCCTGCACCAA 463
Db	81160	TTTGATCTGCTGCTGGGAGTTCCGTCTGTTGAAGTCCATGATGCCACCCCTGCACCA 81101
QY	464	CCTCAGCAACAACGCGAAGCCCCCCCCTATGCTGACATPCCCTCAGAGCCCATGTAGT 523
Db	81100	CCTCAGCAACAACGCGAAGCCCCCCCCTATGCTGACATPCCCTCAGAGCCCTGCAGC 81041
QY	524	CTACCATTCATACCATTCAGTTGATTCAGCACACCGACGCTCTTCGCAACCTTATTC 583
Db	81040	CTACCATTCATACCATTCAGTTGATTCAGCACACCGACGCTCTGCGCAGCCCTTATTC 80981
QY	584	ACAGCTCAGCGCCCAATCAGCAGACAGACAGAGTGTAAAACTGAAGAGTGAACCT 643
Db	80980	ACAGCTCAGCGCCCAATCAGCAGACAGACAGAGTGTAAAACTGAAGAGTGAACCG 80921
QY	644	CTTCCCTGCTGCTGGGTCACTCTCTCCCTGATGACCTTCGCTTTAGATGTAAG 703
Db	80920	CTTCCCTGCTGCTGGGTCACTCTCTCCCTGATGACCTTCGCTTTAGATGTAAG 80861
QY	704	AATCCCATGCACCATTCAGATCCGACAGTACCAGAGAGTGACTTTATTCG 759
Db	80860	AATCCCATGCACCATTCAGATCCGACAGTACCAGAGAGTGACTTTATTCG 80805
RESULT 13		
AC096698/c		
LOCUS	AC096698.7	GI:22218393
DEFINITION	Rattus norvegicus clone CH230-154C12.	*** SEQUENCING IN PROGRESS
ACCESSION	AC096698	
VERSION	AC096698.7	GI:22218393
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS	Muzny,D.,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allien,H., Alsbrooks,S., Amin,A., Angiano,D., Anylebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhaý,C., Burck,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,X., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,J., Haviak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowls,C., Kraft,C.L., Lebow,H., Leván,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,	
	1 (bases 1 to 145971)	
	Muzny,D.,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allien,H., Alsbrooks,S., Amin,A., Angiano,D., Anylebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhaý,C., Burck,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,X., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,J., Haviak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowls,C., Kraft,C.L., Lebow,H., Leván,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,	
	Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mahoney,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olanrunsgboon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D., Snead,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tinney,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villagana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Woodén,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.	
	Unpublished 2 (bases 1 to 145971) Worley,K.C. Direct Submission Submitted (23-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 145971) Rat Genome Sequencing Consortium. Direct Submission Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Aug 14, 2002 this sequence version replaced gi:21735321. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GHMS Center clone name: CH230-154C12 ----- Summary Statistics Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 75503 bases at least Q40 Consensus quality: 79139 bases at least Q30 Consensus quality: 81960 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently consists of 68 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 1518: contig of 1518 bp in length * 1519 1618: gap of unknown length * 1619 3242: contig of 1624 bp in length * 3243 3342: gap of unknown length * 3343 4368: contig of 1026 bp in length * 4369 4468: gap of unknown length * 4469 5700: contig of 1232 bp in length * 5701 5800: gap of unknown length	

* 138287	138386:	gap of unknown length
Query Match	20.6%;	Score 352; DB 2; Length 155250;
Best Local Similarity	90.4%;	Pred. No. 6e-89;
Matches 376;	Conservative	0; Mismatches 40; Indels 0; Gaps 0;
QY	344	AATCTGCAAGACTACGGGAGAGATACCATTATCATCAACCCAGACAGAAAGTTCC 403
Db	81220	AAAATGCTGAGTACTGGGAGAGATCCCATTCATCAGCACAGACAAGAAGTTCC 81161
QY	404	TTCGATTCTCCCAACCGGAGTTCCGTCTGTTGAAGTCCATGATGCCACCCCTGCACCAA 463
Db	81160	TTTGATCTGCTGCTGGGAGTTCCGTCTGTTGAAGTCCATGATGCCACCCCTGCACCA 81101
QY	464	CCTCAGCAACAACGCGAAGCCCCCCCCTATGCTGACATPCCCTCAGAGCCCATGTAGT 523
Db	81100	CCTCAGCAACAACGCGAAGCCCCCCCCTATGCTGACATPCCCTCAGAGCCCTGCAGC 81041
QY	524	CTACCATCATTACCATTCAGTTGATTCAGCACACCGACGCTCTTCGCAACCTTATGCC 583
Db	81040	CTACCATCATTACCATTCAGTTGATTCAGCACACCGACGCTCTGCGCAGCCCTTATGCC 80981
QY	584	ACAGCTCAGCGCCCAATCAGCAGCAGACAGAGTGTAAAACTGAAGAGTGAACCT 643
Db	80980	ACAGCTCAGCGCCCAATCAGCAGCAGACAGAGTGTAAAACTGAAGAGTGAACCG 80921
QY	644	CTTCCCTGCTGCTGGGTCACTCTCCTCCTGATGACCTTCGCTTTAGATGTAAG 703
Db	80920	CTTCCCTGCTGCTGGGTCACTCTCCTCCTGATGACCTTCGCTTTAGATGTAAG 80861
QY	704	AATCCCATGACCATTCAGATCCGACAGTACCAGAGAGTGACTTTATCG 759
Db	80860	AATCCCATGACCATTCAGATCCGACAGTACCAGAGAGTGACTTTATCG 80805
RESULT 13		
AC096698/c		
LOCUS	AC096698.7	GI:22218393
DEFINITION	Rattus norvegicus clone CH230-154C12.	*** SEQUENCING IN PROGRESS
ACCESSION	AC096698	
VERSION	AC096698.7	GI:22218393
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS	Muzny,D.,Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allien,H., Alsbrooks,S., Amin,A., Angiano,D., Anylebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhaý,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,J., Haviak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowls,C., Kraft,C.L., Lebow,H., Leván,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,	
	Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathewey,S., McLeod,M., McNeill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemeleh,O., Okwuonu,G., Olanrunsgboon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D., Snead,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tinney,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villagana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Woodén,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.	
	Unpublished 2 (bases 1 to 145971) Worley,K.C. Direct Submission Submitted (23-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 145971) Rat Genome Sequencing Consortium. Direct Submission Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Aug 14, 2002 this sequence version replaced gi:21735321. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GHMS Center clone name: CH230-154C12 ----- Summary Statistics Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 75503 bases at least Q40 Consensus quality: 79139 bases at least Q30 Consensus quality: 81960 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently consists of 68 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 1518: contig of 1518 bp in length * 1519 1618: gap of unknown length * 1619 3242: contig of 1624 bp in length * 3243 3342: gap of unknown length * 3343 4368: contig of 1026 bp in length * 4369 4468: gap of unknown length * 4469 5700: contig of 1232 bp in length * 5701 5800: gap of unknown length	



Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lohensuhewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Manquim, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, K., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarunpunsagoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 145971)  
Worley, K.C.

Direct Submission  
Submitted (23-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 145971)

REFERENCE  
TITLES  
JOURNAL  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Aug 14, 2002 this sequence version replaced gi:21735321.

COMMENT  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GHMS  
Center clone name: CH230-154C12  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 75503 bases at least Q40  
Consensus quality: 79139 bases at least Q30  
Consensus quality: 81960 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 68 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1518: contig of 1518 bp in length  
1519: gap of unknown length  
1618: contig of 1624 bp in length  
1619: gap of unknown length  
3242: contig of 1026 bp in length  
3243: gap of unknown length  
3343: contig of 1026 bp in length  
3343: gap of unknown length  
4368: contig of 1026 bp in length  
4369: gap of unknown length  
4469: contig of 1232 bp in length  
5700: gap of unknown length  
5701: contig of 1105 bp in length  
5801: gap of unknown length  
6905: gap of unknown length  
6906: contig of 1183 bp in length  
7006: gap of unknown length  
8188: contig of 1376 bp in length  
8189: gap of unknown length  
8288: contig of 1376 bp in length  
9664: gap of unknown length  
9665: contig of 1073 bp in length  
10837: gap of unknown length  
10838: contig of 1162 bp in length  
12099: gap of unknown length  
12100: contig of 1503 bp in length  
13702: gap of unknown length  
13703: contig of 1316 bp in length  
15118: gap of unknown length  
15119: contig of 1444 bp in length  
15219: gap of unknown length  
16662: contig of 1444 bp in length  
16663: gap of unknown length  
16762: contig of 1389 bp in length  
18151: gap of unknown length  
18152: contig of 1828 bp in length  
20079: gap of unknown length  
20179: gap of unknown length  
21451: contig of 1272 bp in length  
21452: gap of unknown length  
22704: contig of 1153 bp in length  
22705: gap of unknown length  
22804: contig of 1170 bp in length  
23974: gap of unknown length  
24074: contig of 1307 bp in length  
25381: gap of unknown length  
25481: contig of 1650 bp in length  
25482: gap of unknown length  
27131: contig of 1783 bp in length  
27231: gap of unknown length  
27232: contig of 1783 bp in length  
29014: gap of unknown length  
29114: contig of 1612 bp in length  
29115: gap of unknown length  
30726: contig of 1612 bp in length  
30727: gap of unknown length  
30826: contig of 1351 bp in length  
32177: gap of unknown length  
32277: contig of 1092 bp in length  
33369: gap of unknown length  
33469: contig of 1534 bp in length  
35003: gap of unknown length  
35103: contig of 1580 bp in length  
36683: gap of unknown length  
36783: contig of 1372 bp in length  
38155: gap of unknown length  
38255: contig of 1021 bp in length  
39276: gap of unknown length  
39277: contig of 1871 bp in length  
41247: gap of unknown length  
41348: contig of 1334 bp in length  
42681: gap of unknown length  
42781: contig of 1867 bp in length  
44648: gap of unknown length  
44748: contig of 1831 bp in length  
46579: gap of unknown length  
46580: contig of 1000 bp in length  
46680: gap of unknown length  
47779: contig of 1839 bp in length  
49618: gap of unknown length  
51952: contig of 2234 bp in length  
52052: gap of unknown length



```
* 52053 53921: contig of 1869 bp in length
* 53922 54021: gap of unknown length
* 54022 55629: contig of 1648 bp in length
* 55629 55769: gap of unknown length
* 55769 57073: contig of 1304 bp in length
* 57073 57173: gap of unknown length
* 57173 59800: contig of 2627 bp in length
* 59800 61117: contig of 1217 bp in length
* 61117 62821: gap of unknown length
* 62821 62921: contig of 1604 bp in length
* 62921 65273: contig of 2352 bp in length
* 65273 65374: gap of unknown length
* 65374 68170: contig of 2797 bp in length
* 68170 68270: gap of unknown length
* 68270 69664: contig of 1394 bp in length
* 69664 69764: gap of unknown length
* 69764 72023: contig of 2259 bp in length
* 72023 72123: gap of unknown length
* 72123 74541: contig of 2418 bp in length
* 74541 76714: contig of 2073 bp in length
* 76714 76814: gap of unknown length
* 76814 79279: contig of 2465 bp in length
* 79279 79379: gap of unknown length
* 79379 81931: contig of 2552 bp in length
* 81931 82031: gap of unknown length
* 82031 84469: contig of 2438 bp in length
* 84469 84569: gap of unknown length
* 84569 86972: contig of 2403 bp in length
* 86972 86972: contig of 2403 bp in length

Query Match      14.8%; Score 253.2; DB 2: Length 145971;
Best Local Similarity 85.3%; Pred. No. 1.3e-60;
Matches 295; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

QY 1303 AGCTTACCAGCGCTTCAAGTCAGAGGTAATGCTTCTCTCTTTGGAAATCTGGCCCA 1362
Db 93957 ATCTTCCCTCCGTTCTAGCTGAGAGGTAATGCTTCTCTCTCTGGAACCTGGCTCA 94016

QY 1363 TGTGAATATGAGCCTCAAGAAAGTGRAGAGGCAATGCTCTCTGGCATGTGTGCTGGG 1422
Db 94017 TGTGAATATGAGCCTCAAGAAAGTGRAGAGGCAATGCTCTGCAATGCGCTGGTGG 94076

QY 1423 CAGTGATCTCTTCAGAGAGCCTATGTGAGGATGAGTGAAGCTGGGATTCCTCAGAGCCC 1482
Db 94077 CAGTGATCTCTTCAGGACCAATGTGAGGATGAGTGAAGCTGGATCCCTCAGAGCCC 94136

QY 1483 TGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCC 1542
Db 94137 TGAGGACTCAGACAGCAGCTATGTTCCCACTCCACTGATAGCCTCATGGGTCATCCCC 94196

QY 1543 TGTTTCAACAGCGCTCAGAGAGAGGATGAGGAAATATAAAGAAAGAGGGAGAT 1602
Db 94197 TGTTTCAATCAGCGCTCAGAGAGAGGATGAGGAAATTTAACTG---CAGAAGGACAT 94253

QY 1603 GTTTTGTCCAGACCTACTAGACCCCAACAGAAAGGTTTGTATTA 1648
Db 94254 TTTCAATCCAGACCTACAGAGAGCCAGAGAGCCCTTGATGTTAA 94299
```

```
RESULT 15
HS33F3F/C
LOCUS HS33F3F 296 bp DNA linear PRI 18-OCT-1995
DEFINITION H.sapiens CpG island DNA genomic MseI fragment, clone 33f3, forward
read cp933f3.ftie.
ACCESSION Z58264
VERSION Z58264.1 GI:1029495
KEYWORDS CpG island; genomic MseI fragment.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 296)
AUTHORS MacDonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE 2 (bases 1 to 296)
AUTHORS Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
TITLE Purification of CpG islands using a methylated DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
PUBMED 8012384
COMMENT Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
http://www.hmp.mrc.ac.uk/ for details
or contact: biocentre@hmp.mrc.ac.uk.
FEATURES
source
location/Qualifiers
1..296
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="33f3"
/sex="male"
/tissue_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"
BASE COUNT 82 a 78 c 84 g 50 t 2 others
ORIGIN
Query Match 12.5%; Score 214.6; DB 9: Length 296;
Best Local Similarity 97.4%; Pred. No. 5.6e-50;
Matches 228; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 27 AGGTCCAGACAGCGCCAGCCCGCTCTGCGACGACGAGTGAATAGTGTGCTACCTCT 86
Db 296 AGGTCCAGACAGCGCCAGCCCGCTCTGCGACGACGAGTGAATAGTGTGCTACCTCT 237

QY 87 TGTCTCGGTTTCAGGTCCAGACCTCCCGCTCTTCCGGCTGCCCTGAACGTCAGGCGACCTC 146
Db 236 TGTCTCGGTTTCAGGTCCAGACCTCCCGCTCTTCCGGCTGCCCTGAACGTCAGGCGACCTC 177

QY 147 AGGACCTGTGATTGGGCGCTTGGCGCGGACCGTGCACCGAGAAACCCCTGGAGGAC 206
Db 176 AGGACCTGTGATTGGGCGCTTGGCGCGGACCGTGCACCGAGAAACCCCTGGAGGAC 117

QY 207 TTGGGCATTCCTTGGGCTCCGCTGCTGT -CTTGGTCTCTCTTTCGGGCAAGTA 259
Db 116 TTGGGCATTCCTTGGGCTCCGCTGCTGTTCCTTCTCTCTTTCGGGCAAGTA 63
```

Search completed: July 18, 2003, 09:03:50  
Job time : 4501 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:03:57 ; Search time 0.492891 Seconds  
(without alignments)  
673.193 Million cell updates/sec

Title: US-09-857-308-5  
Perfect score: 37  
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	33	89.2	413	1	NORB_VIBAL	Q56587 vibrio algi
2	30	81.1	312	1	LPXK_HELPJ	Q9zmb1 helicobacte
3	30	81.1	410	1	NORB_NEIMA	Q9jvp9 neisseria m
4	30	81.1	410	1	NORB_NEIMA	Q9k0m4 neisseria m
5	30	81.1	500	1	GAG_HV1RH	P05890 human immun
6	29	78.4	421	1	GLTT_BACCA	P24944 bacillus ca
7	29	78.4	421	1	GLTT_BACST	P24943 bacillus st
8	29	78.4	492	1	EAAT_ONCVO	Q25605 onchoerca
9	29	78.4	502	1	EA4A_CAEEL	Q22682 caenorhabdi
10	29	78.4	503	1	EA4A_CAEEL	Q10901 caenorhabdi
11	29	78.4	532	1	EA4A_CAEEL	Q21353 caenorhabdi
12	29	78.4	560	1	EA4A_HUMAN	O00341 homo sapien
13	29	78.4	572	1	EA4A_MOUSE	P43006 mus musculu
14	29	78.4	573	1	EA4A_RAT	P31596 rattus norv
15	29	78.4	574	1	EA4A_HUMAN	P43004 homo sapien
16	29	78.4	575	1	EA4A_CAEEL	Q21751 caenorhabdi
17	29	78.4	760	1	SM4A_MOUSE	O62178 mus musculu
18	29	78.4	1197	1	EVGS_ECO57	P58402 escherichia
19	29	78.4	1197	1	EVGS_ECOLI	P30855 escherichia
20	29	78.4	1226	1	YCS3_YEAST	P25357 saccharomyc
21	28	75.7	199	1	AROK_MYCLE	Q9ccs5 mycobacteri
22	28	75.7	233	1	GLP1_MESCR	P45852 mesembryant
23	28	75.7	312	1	LPXK_HELPJ	O25095 helicobacte
24	28	75.7	375	1	AMSC_ERWAM	Q46633 erwinia amy
25	28	75.7	413	1	NORB_VIBHA	Q9rfw0 vibrio harv
26	28	75.7	415	1	NORB_VIBCH	Q9kps2 vibrio chol
27	28	75.7	437	1	GLTP_ECOLI	P21345 escherichia
28	28	75.7	453	1	UCR2_BOVIN	P23004 bos taurus
29	28	75.7	462	1	CTXA_CHIOU	P58762 chiropsalmu
30	28	75.7	469	1	NIFE_SYNPH	O07355 synechococc
31	28	75.7	523	1	EA4A_MOUSE	P51906 mus musculu
32	28	75.7	523	1	EA4A_RAT	P51907 rattus norv
33	28	75.7	524	1	EA4A_BOVIN	Q95135 bos taurus

34 28 75.7 524 1 EAA3\_HUMAN P43005 homo sapien  
35 28 75.7 524 1 EAA3\_RABIT P31597 oryctolagus  
36 28 75.7 578 1 VIPC\_SALTI Q04975 salmonella  
37 28 75.7 825 1 QUTA\_EMENI P10563 emeritella  
38 28 75.7 881 1 DPOL\_RPBGS P03161 ground squi  
39 28 75.7 1377 1 NEOL\_RAT P97603 rattus norv  
40 28 75.7 1443 1 NEOL\_CHICK Q09610 gallus gall  
41 28 75.7 1451 1 SPT6\_YEAST P23615 saccharomyc  
42 28 75.7 1461 1 NEOL\_HUMAN Q92859 homo sapien  
43 28 75.7 1493 1 NEOL\_MOUSE P97798 mus musculu  
44 28 75.7 2210 1 RPPL\_EBOSM O65802 ebola virus  
45 27 73.0 85 1 ITP2\_CHICK Q90683 gallus gall

#### ALIGNMENTS

RESULT 1  
NORB\_VIBAL  
ID NORB\_VIBAL STANDARD; PRT; 413 AA.  
AC Q56587;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)  
DE (Na(+)-translocating NQR subunit B) (Na(+)-NQR subunit B) (NQR complex subunit B) (NQR-1 subunit B).  
GN NQR OR NQR2.  
OS Vibrio alginolyticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=663;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIMB 11038;  
RX MEDLINE=95104445; PubMed=7805867;  
RA Beattie P., Tan K., Bourne R.M., Leach D.R.F., Rich P.R., Ward F.B.;  
RT "Cloning and sequencing of four structural genes for the Na(+)-translocating NADH-ubiquinone oxidoreductase of Vibrio alginolyticus.";  
RT FEBS Lett. 356:333-338(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hayashi M., Unemoto T., Sugiyama A.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-10.  
RX MEDLINE=98149659; PubMed=9490015;  
RA Nakayama Y., Hayashi M., Unemoto T.;  
RT "Identification of six subunits constituting Na+-translocating NADH-quinone reductase from the marine Vibrio alginolyticus.";  
RL FEBS Lett. 422:240-242(1998).  
RN [4]  
RP SEQUENCE OF 1-6 AND 226-237, AND COFACTOR.  
RX MEDLINE=20298664; PubMed=10838078;  
RA Nakayama Y., Yasui M., Sugahara K., Hayashi M., Unemoto T.;  
RT "Covalently bound flavin in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from Vibrio alginolyticus.";  
RL FEBS Lett. 474:165-168(2000).  
RN [5]  
RP INHIBITION OF ENZYMIC ACTIVITY.  
RX MEDLINE=20016049; PubMed=10549856;  
RA Nakayama Y., Hayashi M., Yoshikawa K., Mochida K., Unemoto T.;  
RT "Inhibitor studies of a new antibiotic, koromicin, 2-n-heptyl-4-hydroxyquinoline N-oxide and Ag+ toward the Na+-translocating NADH-quinone reductase from the marine Vibrio alginolyticus.";  
RL Biol. Pharm. Bull. 22:1064-1067(1999).  
RN [6]  
RP COFACTOR, AND MASS SPECTROMETRY.  
RX MEDLINE=21099804; PubMed=11163785;  
RA Hayashi M., Nakayama Y., Yasui M., Maeda M., Furuishi K., Unemoto T.;  
RT "FMN is covalently attached to a threonine residue in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from Vibrio alginolyticus.";

FEBS Lett. 488:5-8(2001).  
 [7]  
 RP REVIEW.  
 RX MEDLINE=21145117; PubMed=11248187;  
 RA Hayashi M., Nakayama Y., Umemoto T.;  
 RT "Recent progress in the Na(+)-translocating NADH-quinone reductase  
 from the marine Vibrio alginolyticus.";  
 RL Biochim. Biophys. Acta 1505:37-44(2001).  
 [8]  
 RN REVIEW.  
 RX MEDLINE=21145118; PubMed=11248188;  
 RA Steuber J.;  
 RT "Na(+)-translocation by bacterial NADH:quinone oxidoreductases: an  
 extension to the complex-I family of primary redox pumps.";  
 RL Biochim. Biophys. Acta 1505:45-56(2001).  
 CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO  
 UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT  
 OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE  
 ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF  
 UBISEMIQUINONE TO UBIQUINOL.  
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +  
 ubiquinol + Na(+)(Out).  
 CC -!- COFACTOR: FMN.  
 CC -!- ENZYME REGULATION: THIS REACTION IS TIGHTLY COUPLED TO THE NA(+)-  
 PUMPING ACTIVITY AND SPECIFICALLY REQUIRES NA(+) FOR ACTIVITY.  
 CC INHIBITED BY KORMICIN AND 2-N-HEPTYL-4-HYDROXYQUINOLINE N-OXIDE  
 (HONO).  
 CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE  
 AND NQRF.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 (potential).  
 CC -!- SIMILARITY: BELONGS TO THE NQRB/RNFD FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; Z37111; CA85477.1; ALT\_INIT.  
 DR EMBL; AB08030; BAA22911.1; -.  
 DR InterPro; IPR004338; NQR2\_Rnfd\_RnFE.  
 DR Pfam; PF03116; NQR2\_Rnfd\_RnFE; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;  
 FT Flavoprotein; FMN; Transmembrane; Inner membrane.  
 FT INIT\_MET 0 0  
 FT TRANSMEM 22 39 POTENTIAL.  
 FT TRANSMEM 59 81 POTENTIAL.  
 FT TRANSMEM 124 146 POTENTIAL.  
 FT TRANSMEM 158 180 POTENTIAL.  
 FT TRANSMEM 267 289 POTENTIAL.  
 FT TRANSMEM 296 315 POTENTIAL.  
 FT TRANSMEM 325 344 POTENTIAL.  
 FT TRANSMEM 357 374 POTENTIAL.  
 FT TRANSMEM 378 400 POTENTIAL.  
 FT TRANSMEM 235 235 FMN.  
 FT BINDING 228 228 A -> V (IN REF. 1).  
 FT CONFLICT 294 294 W -> G (IN REF. 1).  
 FT CONFLICT 294 294  
 SQ SEQUENCE 413 AA; 45211 MW; A6F849F7A5C91C9D CRC64;  
 Query Match 89.2%; Score 33; DB 1; Length 413;  
 Best Local Similarity 75.0%; Pred. No. 4.2;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LYQAVATI 8  
 Db 25 LYEAVATV 32  
 RESULT 2  
 LPXK\_HELPJ

ID LPXK\_HELPJ STANDARD; PRT; 312 AA.  
 AC Q9ZMB1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tetraacyldisaccharide 4'-kinase (EC 2.7.1.130) (Lipid A 4'-kinase).  
 GN LPXK OR JHP0311.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 CC Helicobacter.  
 CC NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 gastric pathogen Helicobacter pylori.";  
 RL Nature 397:176-180(1999).  
 CC -!- FUNCTION: TRANSFERS THE GAMMA-PHOSPHATE OF ATP TO THE 4'-POSITION  
 OF A TETRAACYLDISACCHARIDE 1-PHOSPHATE INTERMEDIATE (TERMED DS-1-  
 P) TO FORM TETRAACYLDISACCHARIDE 1,4'-BIS-PHOSPHATE (LIPID IVA)  
 (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: ATP + 2,3-bis(3-hydroxytetradecanoyl)-D-  
 glucosaminyl-(beta-D-1,6)-2,3-bis(3-hydroxytetradecanoyl)-D-  
 glucosaminyl beta-phosphate = ADP + 2,3,2',3'-tetrakis(3-  
 hydroxytetradecanoyl)-D-glucosaminyl-1,6-beta-D-glucosamine 1,4'-  
 bisphosphate.  
 CC -!- PATHWAY: Lipid A biosynthesis; sixth step.  
 CC -!- SIMILARITY: BELONGS TO THE LPXK FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE001467; AAD05884.1; -.  
 DR InterPro; IPR003758; LpxK.  
 DR Pfam; PF02606; LpxK; 1.  
 DR TIGRFAMS; TIGR00682; lpxk; 1.  
 KW Transferase; Kinase; Lipid A biosynthesis; Lipid synthesis;  
 KW ATP-binding; Complete proteome.  
 FT NP\_BIND 60 67 ATP (POTENTIAL).  
 SQ SEQUENCE 312 AA; 35567 MW; 43E44608F3A60FAA CRC64;  
 Query Match 81.1%; Score 30; DB 1; Length 312;  
 Best Local Similarity 62.5%; Pred. No. 16;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LYQAVATI 8  
 Db 32 LYQCIATI 39  
 RESULT 3  
 NQRB\_NEIMA  
 ID NQRB\_NEIMA STANDARD; PRT; 410 AA.  
 AC Q9JVP9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)  
 DE (Na(+)-translocating NQR subunit B) (Na(+)-NQR subunit B) (NQR complex  
 subunit B) (NQR-1 subunit B).  
 GN NQRB OR NMA0751.  
 OS Neisseria meningitidis (serogroup A).  
 CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

```

OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis z2491.";
RL Nature 404:502-506(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +
CC ubiquinol + Na(+)(Out).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE NQRB/RNFD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162754; CAB84034.1; -
DR InterPro; IPR004338; NQR2_Rnfd_Rnfe.
DR Pfam; PF03116; NQR2_Rnfd_Rnfe; 1.
DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
DR Flavoprotein; FMN; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 58 80 POTENTIAL.
FT TRANSMEM 120 142 POTENTIAL.
FT TRANSMEM 155 177 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 293 312 POTENTIAL.
FT TRANSMEM 322 341 POTENTIAL.
FT TRANSMEM 354 371 POTENTIAL.
FT TRANSMEM 375 397 POTENTIAL.
FT TRANSMEM 232 232 FMN (BY SIMILARITY).
FT BINDING 232 232
SQ SEQUENCE 410 AA; 44536 MW; D46528F211513ECE CRC64;

Query Match 81.1%; Score 30; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LYQAVATI 8
Db 26 LYEAATI 33

RESULT 4
NORB_NEIMB STANDARD; PRT; 410 AA.
AC Q9K0M4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit B) (Na(+)-NQR subunit B) (NQR complex
DE subunit B) (NQR-1 subunit B).
GN NQRB OR NMB0568.
OS Neisseria meningitidis (serogroup B).

```

```

CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiugani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NQRA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +
CC ubiquinol + Na(+)(Out).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE NQRB/RNFD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002412; AAF40996.1; -
DR TIGR; NMB0568; -
DR InterPro; IPR004338; NQR2_Rnfd_Rnfe.
DR Pfam; PF03116; NQR2_Rnfd_Rnfe; 1.
DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
DR Flavoprotein; FMN; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 58 80 POTENTIAL.
FT TRANSMEM 120 142 POTENTIAL.
FT TRANSMEM 155 177 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 293 312 POTENTIAL.
FT TRANSMEM 322 341 POTENTIAL.
FT TRANSMEM 354 371 POTENTIAL.
FT TRANSMEM 375 397 POTENTIAL.
FT TRANSMEM 232 232 FMN (BY SIMILARITY).
FT BINDING 232 232
SQ SEQUENCE 410 AA; 44601 MW; F632E12206170B4F CRC64;

Query Match 81.1%; Score 30; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LYQAVATI 8
Db 26 LYEAATI 33

RESULT 5
GAG_HV1RH STANDARD; PRT; 500 AA.
AC P05890;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAG polyprotein [Contains: Core proteins P17, P24, P2, P7, P1, P6].

```

GN GAG.  
OS Human immunodeficiency virus type 1 (RF/HAT isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11701;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Starch B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S.,  
RA Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C.,  
RA Wong-Staal F.;  
RL Submitted (XXX-1987) to the HIV data bank.  
CC -!- FUNCTION: PERFORMS HIGHLY COMPLEX ORCHESTRATED TASKS DURING THE  
CC ASSEMBLY, BUDDING, MATURATION, AND INFECTION STAGES OF THE VIRAL  
CC REPLICATION CYCLE. DURING VIRAL ASSEMBLY, THE PROTEINS FORM  
CC MEMBRANE ASSOCIATIONS AND SELF-ASSOCIATIONS THAT ULTIMATELY  
CC RESULT IN BUDDING OF AN IMMATURE VIRION FROM THE INFECTED CELL.  
CC GAG PRECURSORS ALSO FUNCTION DURING VIRAL ASSEMBLY TO SELECTIVELY  
CC BIND AND PACKAGE TWO PLUS STRANDS OF GENOMIC RNA.  
CC -!- PTM: THE P24 PROTEIN IS PHOSPHORYLATED.  
CC -!- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M17451; AAA45052.1; -  
CC HSSP: P05888; IAAF.  
CC HIV: M17451; GAG\$RF.  
CC InterPro: IPR000721; Gag\_p24.  
CC InterPro: IPR000071; Retrovir\_p17.  
CC InterPro: IPR001878; Znf\_CCHC.  
CC Pfam: PF00098; zf-CCHC; 2.  
CC Pfam: PF00540; Gag\_p17; 1.  
CC Pfam: PF00607; Gag\_p24; 1.  
CC PRINTS: PR00939; C2HCZNFINGER.  
CC PRINTS: PR00234; HIVIMATRIX.  
CC SMART: SM00343; Znf\_C2HC; 2.  
CC PROSITE: PS0158; ZF\_CCHC; 2.  
KW AIDS: Core protein; Polyprotein; Myristate; Phosphorylation;  
KW Zinc-finger; Repeat.  
FT INIT\_MET 0 BY SIMILARITY.  
FT CHAIN 1 131 CORE PROTEIN P17 (MATRIX PROTEIN).  
FT CHAIN 132 362 CORE PROTEIN P24 (CORE ANTIGEN).  
FT CHAIN 363 376 CORE PROTEIN P2.  
FT CHAIN 377 433 CORE PROTEIN P7 (NUCLEOCAPSID PROTEIN).  
FT CHAIN 434 449 CORE PROTEIN P1.  
FT CHAIN 450 500 CORE PROTEIN P6.  
FT ZN\_FING 389 406 CCHC-TYPE 1.  
FT ZN\_FING 410 427 CCHC-TYPE 2.  
FT LIPID 1 MYRISTATE (BY SIMILARITY).  
SQ SEQUENCE 500 AA; 55825 MW; 44AA0CB5CD9EFF7B CRC64;  
  
Query Match 81.1%; Score 30; DB 1; Length 500;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LYQAVATI 8  
Db 77 LYNVAATL 84  
  
RESULT 6  
GLTT\_BACCA  
ID GLTT\_BACCA STANDARD; PRT; 421 AA.  
AC P24944;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Proton/sodium-glutamate symport protein (Glutamate-aspartate carrier  
DE protein).

GN GLTT.  
OS Bacillus caldotenax.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1395;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=93062018; PubMed=1359385;  
RA Tolner B., Poolman B., Konings W.N.;  
RT "Characterization and functional expression in Escherichia coli of  
RT the sodium/proton/glutamate symport proteins of Bacillus  
RT stearothermophilus and Bacillus caldotenax.";  
RL Mol. Microbiol. 6:2845-2856(1992).  
CC -!- FUNCTION: THIS CARRIER PROTEIN IS PART OF THE NA(+)-DEPENDENT,  
CC BINDING-PROTEIN-INDEPENDENT GLUTAMATE-ASPARTATE TRANSPORT SYSTEM.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE SODIUM-DICARBOXYLATE SYMPORTER FAMILY  
CC (SDF, TC 2.A.23).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M86509; AAA22493.1; -  
CC PIR: S26246; S26246.  
CC InterPro: IPR001991; Na/dico\_symp.  
CC Pfam: PF00375; SDF; 1.  
CC PROSITE: PS00713; NA-DICARBOXYL\_SYMP\_1; 1.  
CC PROSITE: PS00714; NA-DICARBOXYL\_SYMP\_2; 1.  
KW Transport; Transmembrane; Symport.  
FT DOMAIN 1 3 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 4 24 POTENTIAL.  
FT DOMAIN 25 43 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 44 64 POTENTIAL.  
FT DOMAIN 65 77 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 78 98 POTENTIAL.  
FT DOMAIN 99 148 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 149 169 POTENTIAL.  
FT DOMAIN 170 198 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 199 219 POTENTIAL.  
FT DOMAIN 220 222 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 223 243 POTENTIAL.  
FT DOMAIN 244 244 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 245 265 POTENTIAL.  
FT DOMAIN 266 306 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 307 327 POTENTIAL.  
FT DOMAIN 328 330 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 331 351 POTENTIAL.  
FT TRANSMEM 352 372 POTENTIAL.  
FT DOMAIN 373 421 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 421 AA; 45345 MW; 10ABBE12EDD1E7E4 CRC64;  
  
Query Match 78.4%; Score 29; DB 1; Length 421;  
Best Local Similarity 75.0%; Pred. No. 37;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LYQAVATI 8  
Db 308 LYQALAAI 315  
  
RESULT 7  
GLTT\_BACST  
ID GLTT\_BACST STANDARD; PRT; 421 AA.  
AC P24943;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Proton/sodium-glutamate symport protein (Glutamate-aspartate carrier  
DE protein).

```
GN GLTT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 7954;
RX MEDLINE=93062018; PubMed=1359385;
RA Tolner B., Poolman B., Konings W.N.;
RT "Characterization and functional expression in Escherichia coli of
the sodium/proton/glutamate symport proteins of Bacillus
stearothermophilus and Bacillus caldotenax.";
RL Mol. Microbiol. 6:2845-2856(1992).
CC -1- FUNCTION: THIS CARRIER PROTEIN IS PART OF THE NA(+)-DEPENDENT,
BINDING-PROTEIN-INDEPENDENT GLUTAMATE-ASPARTATE TRANSPORT SYSTEM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
(SDF, TC 2.A.23).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86508; AAA22492.1; .
DR PIR; S26247; S26247.
DR InterPro; IPR001991; Na/dico_sympt.
DR Pfam; PF00375; SDF; 1.
DR PROSITE; PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE; PS00714; NA_DICARBOXYL_SYMPT_2; 1.
DR Transport; Transmembrane; Symport.
KW DOMAIN 1 3 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 4 24 POTENTIAL.
FT DOMAIN 25 43 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 44 64 POTENTIAL.
FT DOMAIN 65 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 98 POTENTIAL.
FT DOMAIN 99 148 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 149 169 POTENTIAL.
FT DOMAIN 170 198 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 199 219 POTENTIAL.
FT DOMAIN 220 222 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 223 243 POTENTIAL.
FT DOMAIN 244 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 265 POTENTIAL.
FT DOMAIN 266 306 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 307 327 POTENTIAL.
FT DOMAIN 328 330 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT DOMAIN 373 421 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 421 AA; 45469 MW; 30C5738E8FD3A54F CRC64;

Query Match 78.4%; Score 29; DB 1; Length 421;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 308 LYQALAAI 315

RESULT 8
ID EAAT_ONCVO STANDARD; PRT; 492 AA.
AC Q25605;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Excitatory amino acid transporter (Sodium-dependent glutamate/
```

```
DE aspartate transporter).
GN GLT-1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97039669; PubMed=8885221;
RA Radice A.D., Lustigman S.;
RT "Cloning and characterization of cDNAs encoding putative glutamate
transporters from Caenorhabditis elegans and Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 80:41-53(1996).
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
AS A SYMPORT BY CO-TRANSPORTING SODIUM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
(SDF, TC 2.A.23).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U35251; AAB41937.1; .
DR InterPro; IPR001991; Na/dico_sympt.
DR Pfam; PF00375; SDF; 1.
DR PRINTS; PR00173; EDTRNSPORT.
DR PROSITE; PS00713; NA_DICARBOXYL_SYMPT_1; 1.
DR PROSITE; PS00714; NA_DICARBOXYL_SYMPT_2; 1.
DR Transport; Transmembrane; Glycoprotein; Symport.
KW DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT DOMAIN 106 191 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 228 248 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 358 378 POTENTIAL.
FT TRANSMEM 389 409 POTENTIAL.
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 492 AA; 53391 MW; CEE52D670F76A89E CRC64;

Query Match 78.4%; Score 29; DB 1; Length 492;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8
Db 355 LYEAAAI 362

RESULT 9
ID EAA4_CAEEL STANDARD; PRT; 502 AA.
AC Q22682;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative sodium-dependent excitatory amino acid transporter T22E5.2.
GN T22E5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
```

RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Minx P.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -|- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
CC (SDF, TC 2.A.23).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U43282; AAA83614.1; -;  
DR WormPep: T22E5.2; CE04992.  
DR InterPro: IPR001991; Na/dico\_symp.  
DR Pfam: PF00375; SDF; 1.  
DR PRINTS: PR00173; EDTNSPORT.  
DR PROSITE: PS00713; NA\_DICARBOXYL\_SYMP\_1; 1.  
DR PROSITE: PS00714; NA\_DICARBOXYL\_SYMP\_2; 1.  
KW Hypothetical protein; Transport; transmembrane; Glycoprotein; Symport.  
FT DOMAIN 1 31 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 32 52 POTENTIAL.  
FT TRANSMEM 74 94 POTENTIAL.  
FT TRANSMEM 111 131 POTENTIAL.  
FT DOMAIN 132 217 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 218 238 POTENTIAL.  
FT TRANSMEM 258 278 POTENTIAL.  
FT TRANSMEM 300 320 POTENTIAL.  
FT TRANSMEM 388 408 POTENTIAL.  
FT TRANSMEM 419 439 POTENTIAL.  
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 502 AA; 54610 MW; 8B4C7469C594727D CRC64;  
  
Query Match 78.4%; Score 29; DB 1; Length 502;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 LYQAVATI 8  
II:III I  
Db 385 LYEAVAAI 392  
  
RESULT 10  
EAAI\_CAEEL STANDARD; PRT; 503 AA.  
AC Q10901; P90798; Q17920;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Excitatory amino acid transporter (Sodium-dependent glutamate/  
DE aspartate transporter).  
GN Glut-1 or c12d12.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RX MEDLINE=97039669; PubMed=8885221;  
RA Radice A.D., Lustigman S.;  
RT "Cloning and characterization of cDNAs encoding putative glutamate  
RT transporters from Caenorhabditis elegans and Onchocerca volvulus.";  
RL Mol. Biochem. Parasitol. 80:41-53(1996).  
[2]  
RN SEQUENCE FROM N.A. (SHORT FORM).  
RC STRAIN-Bristol N2;  
RX MEDLINE=97079197; PubMed=8920929;  
RA Kawano T., Takuwa K., Nakajima T.;

RT "Molecular cloning of a cDNA for the glutamate transporter of the  
RL nematode Caenorhabditis elegans.";  
RL Biochem. Biophys. Res. Commun. 228:415-420(1996).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN-Bristol N2;  
RA Nhan M., Hawkins J.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
[4]  
RN REVISIONS.  
RP Waterston R.;  
RA Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
[5]  
RN SEQUENCE FROM N.A.  
RP STRAIN-Bristol N2;  
RC MEDLINE=97321871; PubMed=9178573;  
RX Kawano T., Takuwa K., Nakajima T.;  
RT "Structure and activity of a new form of the glutamate transporter of  
RL the nematode Caenorhabditis elegans.";  
RL Biosci. Biotechnol. Biochem. 61:927-929(1997).  
CC -|- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.  
CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY  
CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS  
CC AS A SYMPORT BY CO-TRANSPORTING SODIUM (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GLT-1 AND GLT-2 (SHOWN HERE);  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -|- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
CC (SDF, TC 2.A.23).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U35250; AAB41909.1; -;  
DR EMBL: U35250; AAB41910.1; -;  
DR EMBL: D86740; BAA13164.1; -;  
DR EMBL: U51998; AAL00857.1; -;  
DR EMBL: U51998; AAL00858.1; -;  
DR EMBL: D86741; BAA21840.1; -;  
DR WormPep: C12D12.2a; CE29083.  
DR WormPep: C12D12.2b; CE29084.  
DR InterPro: IPR001991; Na/dico\_symp.  
DR Pfam: PF00375; SDF; 1.  
DR PRINTS: PR00173; EDTNSPORT.  
DR PROSITE: PS00713; NA\_DICARBOXYL\_SYMP\_1; 1.  
DR PROSITE: PS00714; NA\_DICARBOXYL\_SYMP\_2; 1.  
KW Transport; Transmembrane; Glycoprotein; Symport; Alternative splicing.  
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 19 39 POTENTIAL.  
FT TRANSMEM 59 79 POTENTIAL.  
FT TRANSMEM 96 116 POTENTIAL.  
FT DOMAIN 117 198 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 199 219 POTENTIAL.  
FT TRANSMEM 239 259 POTENTIAL.  
FT TRANSMEM 281 301 POTENTIAL.  
FT TRANSMEM 369 389 POTENTIAL.  
FT TRANSMEM 400 420 POTENTIAL.  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 1 11 MISSING (IN ISOFORM GLT-1).  
FT CONFLICT 68 68 L -> P (IN REF. 1).  
FT CONFLICT 273 273 ADTA -> ENTT (IN REF. 1).  
FT CONFLICT 321 321 A -> G (IN REF. 1).  
FT CONFLICT 480 503 AMNDEKRLAVYNSLPTDDEKHTH -> GFVF (IN  
FT REF. 3).  
SQ SEQUENCE 503 AA; 54675 MW; 3014BFF28E41E798 CRC64;  
  
Query Match 78.4%; Score 29; DB 1; Length 503;

Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 366 LYEAVAI 373

RESULT 11  
EAA2\_CAEEL STANDARD; PRT; 532 AA.  
AC Q21353;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative sodium-dependent excitatory amino acid transporter K08F4.4.  
GN K08F4.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Hendry C.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
(SDF, TC 2.A.23).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z68879; CAA93084.1; -;  
DR WormPep; K08F4.4; CE06152.  
DR InterPro; IPR001991; Na/dico\_symp.  
DR Pfam; PF00375; SDF; 1.  
DR PRINTS; PR00173; EDRNSPORT.  
DR PROSITE; PS00713; NA\_DICARBOXYL\_SYMPT\_1; 1.  
DR PROSITE; PS00714; NA\_DICARBOXYL\_SYMPT\_2; 1.  
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport.  
FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 6 26 POTENTIAL.  
FT TRANSMEM 46 66 POTENTIAL.  
FT TRANSMEM 83 103 POTENTIAL.  
FT DOMAIN 104 181 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 182 202 POTENTIAL.  
FT TRANSMEM 222 242 POTENTIAL.  
FT TRANSMEM 264 284 POTENTIAL.  
FT TRANSMEM 352 372 POTENTIAL.  
FT TRANSMEM 383 402 POTENTIAL.  
FT CARBOHYD 164 164 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 169 169 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 532 AA; 58469 MW; FA370EAD3795E5DF CRC64;

Query Match 78.4%; Score 29; DB 1; Length 532;  
Best Local Similarity 75.0%; Pred. No. 47;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 349 LYEAVAI 356

RESULT 12  
EAA5\_HUMAN STANDARD; PRT; 560 AA.  
ID EAA5\_HUMAN PRT; 560 AA.  
AC O00341;

DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Excitatory amino acid transporter 5 (Retinal glutamate transporter).  
GN SLCL1A7 OR EAAT5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RX MEDLINE=97268713; PubMed=9108121;  
RA Arriza J.L., Eliasof S., Kavanaugh M.P., Amara S.G.;  
RT "Excitatory amino acid transporter 5, a retinal glutamate transporter  
RT coupled to a chloride conductance.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:4155-4160(1997).  
CC -!- FUNCTION: TRANSPORTS L-GLUTAMATE; THE L-GLUTAMATE UPTAKE IS  
CC SODIUM- AND VOLTAGE-DEPENDENT AND CHLORIDE-INDEPENDENT. THE EAAT5-  
CC ASSOCIATED CHLORIDE CONDUCTANCE MAY PARTICIPATE IN VISUAL  
CC PROCESSING.  
CC -!- SUBUNIT: INTERACTS WITH THE PDZ DOMAINS OF DLG4.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN RETINA. DETECTABLE IN  
CC LIVER, HEART, MUSCLE AND BRAIN.  
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
(SDF, TC 2.A.23).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U76362; AAB53971.1; -;  
DR Genew; HGNC:10945; SLC1A7.  
DR MIN; 604471; -;  
DR InterPro; IPR001991; Na/dico\_symp.  
DR Pfam; PF00375; SDF; 1.  
DR PRINTS; PR00173; EDRNSPORT.  
DR PROSITE; PS00713; NA\_DICARBOXYL\_SYMPT\_1; 1.  
DR PROSITE; PS00714; NA\_DICARBOXYL\_SYMPT\_2; 1.  
KW Transport; Transmembrane; Glycoprotein; Symport; Multigene family.  
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 17 37 POTENTIAL.  
FT TRANSMEM 60 80 POTENTIAL.  
FT TRANSMEM 94 114 POTENTIAL.  
FT DOMAIN 115 216 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 217 237 POTENTIAL.  
FT TRANSMEM 260 280 POTENTIAL.  
FT TRANSMEM 300 320 POTENTIAL.  
FT TRANSMEM 330 350 POTENTIAL.  
FT TRANSMEM 372 392 POTENTIAL.  
FT TRANSMEM 414 434 POTENTIAL.  
FT TRANSMEM 457 477 POTENTIAL.  
FT CARBOHYD 191 191 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 560 AA; 60722 MW; C71D48355AACFF32 CRC64;

Query Match 78.4%; Score 29; DB 1; Length 560;  
Best Local Similarity 75.0%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 383 LYEAVAI 390

RESULT 13  
EAA2\_MOUSE  
ID EAA2\_MOUSE STANDARD; PRT; 572 AA.  
AC P43006; O35877; O54686; O54687;

DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Excitatory amino acid transporter 2 (Sodium-dependent  
 DE glutamate/aspartate transporter 2) (GLT-1).  
 GN SLC1A2 OR EAAT2 OR GLT1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=95213010; PubMed=7698742;  
 RA Kirschner M.A., Copeland N.G., Gilbert D.J., Jenkins N.A.,  
 RA Anara S.G.;  
 RT "Mouse excitatory amino acid transporter EAAT2: isolation,  
 RT characterization, and proximity to neuroexcitability loci on mouse  
 RT chromosome 2.";   
 RL Genomics 24:218-224(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCL:ICR; TISSUE=Cerebellum;  
 RX MEDLINE=95284091; PubMed=7766684;  
 RA Mukainaka Y., Tanaka K., Hagiwara T., Wada K.;  
 RT "Molecular cloning of two glutamate transporter subtypes from mouse  
 RT brain.";   
 RL Biochim. Biophys. Acta 1244:233-237(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=96032356; PubMed=7557442;  
 RA Sutherland M.L., Delaney T.A., Noebels J.L.;  
 RT "Molecular characterization of a high-affinity mouse glutamate  
 RT transporter.";   
 RL Gene 162:271-274(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Peng J.-B., Guo L.-H.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=JCL:ICR; TISSUE=Brain, and Liver;  
 RX MEDLINE=98039013; PubMed=9373176;  
 RA Utsumiya-Tate N., Endou H., Kanai Y.;  
 RT "Tissue specific variants of glutamate transporter GLT-1.";   
 RL FEBS Lett. 416:312-316(1997).  
 CC -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.  
 CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY  
 CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS  
 CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GLT-1 (SHOWN HERE), GLT-1A  
 CC AND GLT-1B; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: ISOFORM GLT1 IS EXPRESSED IN THE BRAIN.  
 CC ISOFORMS GLT-1A AND GLT-1B ARE EXPRESSED IN THE LIVER.  
 CC -!- PTM: GLYCOSYLATED.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
 CC (SDF, TC 2.A.23).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; U11763; AAA77673.1; -;  
 DR EMBL; D43796; BAA07854.1; -;  
 DR EMBL; U24699; AAA91643.1; -;  
 DR EMBL; U75372; AAB71737.1; -;

DR EMBL; U75373; AAB71738.1; -;  
 DR EMBL; AB007810; BAA23770.1; -;  
 DR EMBL; AB007811; BAA23771.1; -;  
 DR EMBL; AB007812; BAA23772.1; -;  
 DR MGD; MGI:101931; Slc1a2.  
 DR InterPro: IPR001991; Na/dico\_symp.  
 DR Pfam: PF00375; SDF; 1.  
 DR PRINTS: PR00173; EDTRNSPORT.  
 DR PROSITE: PS00713; NA-DICARBOXYL\_SYMP\_1; 1.  
 DR PROSITE: PS00714; NA-DICARBOXYL\_SYMP\_2; 1.  
 KW Transport; transmembrane; Glycoprotein; Symport; Multigene family;  
 KW Alternative splicing.  
 FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 45 64 POTENTIAL.  
 FT TRANSMEM 88 108 POTENTIAL.  
 FT TRANSMEM 121 142 POTENTIAL.  
 FT DOMAIN 143 238 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 239 258 POTENTIAL.  
 FT TRANSMEM 279 300 POTENTIAL.  
 FT TRANSMEM 316 338 POTENTIAL.  
 FT TRANSMEM 405 429 POTENTIAL.  
 FT TRANSMEM 436 458 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 215 215 MASTEG -> MVS (IN ISOFORMS GLT-1A AND  
 FT VARSPLIC 1 6 ISOFORM GLT-1B).  
 FT VARSPLIC 551 572 TLAANGKSDCSVEEPPKREK -> PPFPLDIETCI  
 FT CONFLICT 26 26 (IN ISOFORM GLT-1B).  
 FT CONFLICT 62 62 D -> E (IN REF. 3).  
 FT CONFLICT 112 112 G -> R (IN REF. 3).  
 FT CONFLICT 454 454 A -> V (IN REF. 3).  
 FT CONFLICT 525 525 T -> I (IN REF. 4).  
 FT CONFLICT 572 572 K -> L (IN REF. 4).  
 FT CONFLICT 572 572 K -> EFD (IN REF. 3).  
 SQ SEQUENCE 572 AA; 62030 MW; 13C7C30DED40CA81 CRC64;  
 Query Match 78.4%; Score 29; DB 1; Length 572;  
 Best Local Similarity 75.0%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LYQAVATI 8  
 Db 402 LYEAVAAI 409  
 RESULT 14  
 EAA2\_RAT  
 ID EAA2\_RAT STANDARD; PRT; 573 AA.  
 AC P31596;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Excitatory amino acid transporter 2 (Sodium-dependent  
 DE glutamate/aspartate transporter 2) (GLUT-R) (GLT-1).  
 GN SLC1A2 OR EAAT2 OR GLT1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93078876; PubMed=1448170;  
 RA Pines G., Danbolt N.C., Bjoeraas M., Zhang Y., Bendahan A., Eide L.,  
 RA Koepsell H., Storm-Mathisen J., Seeborg E., Kanner B.I.;  
 RT "Cloning and expression of a rat brain L-glutamate transporter.";   
 RL Nature 360:464-467(1992).  
 RN [2]  
 RP ERRATUM.  
 RA Pines G., Danbolt N.C., Bjoeraas M., Zhang Y., Bendahan A., Eide L.,  
 RA Koepsell H., Storm-Mathisen J., Seeborg E., Kanner B.I.;  
 RL Nature 360:768-768(1992).  
 RN [3]



RP REVISIONS TO 260-289.  
RX MEDLINE=93292859; PubMed=8099882;  
RA Kanner B.I.;  
RT "Glutamate transporters from brain. A novel neurotransmitter  
transporter family";  
RL FEBS Lett. 325:95-99(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Forebrain;  
RA Roginski R.S., Choudhury K., Meiners S., Marone M., Basma A.N.,  
Geller H.M.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM GLT-1A).  
RC TISSUE=Hepatoma;  
RA Pollard M., McGivan J.;  
RX MEDLINE=20521659; PubMed=11068035;  
RT "The rat hepatoma cell line H4-II-E-C3 expresses high activities of  
the high-affinity glutamate transporter GLT-1A";  
RL FEBS Lett. 484:74-76(2000).  
RN [6]  
RP MUTAGENESIS OF LYS-298 AND HIS-326.  
RX MEDLINE=94308247; PubMed=7913472;  
RA Zhang Y., Pines G., Kanner B.I.;  
RT "Histidine 326 is critical for the function of GLT-1, a (Na<sup>+</sup> + K<sup>+</sup>)-  
coupled glutamate transporter from rat brain";  
RL J. Biol. Chem. 269:19573-19577(1994).  
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.  
ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY  
RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS  
AS A SYMPORT BY CO-TRANSPORTING SODIUM.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GLT1 (SHOWN HERE) AND GLT-1A;  
ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: LOCALIZED IN BRAIN AND IS HIGHLY ENRICHED IN  
THE PURKINJE CELL LAYER IN CEREBELLUM.  
CC -1- PTM: GLYCOSYLATED.  
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
(SDF, TC 2.A.23).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X67857; CAA48042.1; ALT\_SEQ.  
DR EMBL; U15098; AAA93061.1; -  
DR EMBL; U15098; AAA93062.1; ALT\_INIT.  
DR EMBL; AF297648; AAG13411.1; -  
DR InterPro: IPR001991; Na/dico\_symp.  
DR Pfam: PF00375; SDF; 1.  
DR PRINTS: PR00173; EDPRNSPORT.  
DR PROSITE: PS00713; NA\_DICARBOXYL\_SYMPT\_1; 1.  
DR PROSITE: PS00714; NA\_DICARBOXYL\_SYMPT\_2; 1.  
KW Transport; Transmembrane; Glycoprotein; Symport; Multigene family;  
KW Alternative splicing.  
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 45 64 POTENTIAL.  
FT TRANSMEM 88 108 POTENTIAL.  
FT TRANSMEM 121 142 POTENTIAL.  
FT DOMAIN 143 238 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 239 258 POTENTIAL.  
FT TRANSMEM 279 300 POTENTIAL.  
FT TRANSMEM 316 338 POTENTIAL.  
FT TRANSMEM 405 429 POTENTIAL.  
FT TRANSMEM 436 458 POTENTIAL.  
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPPLIC 1 6 MASTEP -> MVS (IN ISOFORM GLT-1A).  
FT MUTAGEN 298 K->N,T: REDUCED TRANSPORTER ACTIVITY.

FT MUTAGEN 298 K->H,R: NORMAL TRANSPORTER ACTIVITY.  
FT MUTAGEN 326 H->N,T,K,R: NO TRANSPORTER ACTIVITY.  
FT CONFLICT 521 V -> I (IN REF. 4).  
SQ SEQUENCE 573 AA; 62106 MW; 8C51D30954E00E7F CRC64;  
Query Match 78.4%; Score 29; DB 1; Length 573;  
Best Local Similarity 75.0%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 LYQAVATI 8  
DB 402 LYEAVAAI 409  
RESULT 15  
EAA2\_HUMAN  
ID EAA2\_HUMAN STANDARD; PRT; 574 AA.  
AC P43004; Q14417;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Excitatory amino acid transporter 2 (Sodium-dependent  
glutamate/aspartate transporter 2).  
GN SLC1A2 OR EAAT2 OR GLT1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain Cortex;  
RX MEDLINE=94365697; PubMed=7521911;  
RA Arriza J.L., Fairman W.A., Wendy A., Wadiche J.I., Murdoch G.H.,  
Kavanaugh M.P., Amara S.G.;  
RT "Functional comparisons of three glutamate transporter subtypes  
cloned from human motor cortex";  
RL J. Neurosci. 14:5559-5569(1994).  
RN [2]  
RP SEQUENCE OF 11-574 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=94227088; PubMed=8172925;  
RA Shashidharan P., Wittenberg I., Plaitakis A.;  
RT "Molecular cloning of human brain glutamate/aspartate transporter  
II";  
RL Biochim. Biophys. Acta 1191:393-396(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=95002073; PubMed=7522567;  
RA Manfras B.J., Rudert W.A., Trucco M., Boehm B.O.;  
RT "Cloning and characterization of a glutamate transporter cDNA from  
human brain and pancreas";  
RL Biochim. Biophys. Acta 1195:185-188(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=98130689; PubMed=9463476;  
RA Shimamoto K., Lebrun B., Yasuda-Kamatani Y., Sakaitani M., Shigeri Y.,  
Yumoto N., Nakajima T.;  
RT "DL-thereo-b-benzoyloxyaspartate, a potent blocker of excitatory amino  
acid transporters";  
RL Mol. Pharmacol. 53:195-201(1998).  
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.  
ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY  
RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS  
AS A SYMPORT BY CO-TRANSPORTING SODIUM.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- PTM: GLYCOSYLATED.  
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
(SDF, TC 2.A.23).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U03505; AAA50429.1; -;  
DR EMBL; U01824; AAA18900.1; -;  
DR EMBL; Z32517; CAA83532.1; -;  
DR EMBL; D85884; BAA28706.1; -;  
DR Genew; HGNC:10940; SLC1A2.  
DR MIM; 600300; -;  
DR InterPro; IPR001991; Na/dico\_symp.  
DR Pfam; PF00375; SDF; 1.  
DR PRINTS; PR00173; EDRNSPORT.  
DR PROSITE; PS00713; NA\_DICARBOXYL\_SYMP.1; 1.  
DR PROSITE; PS00714; NA\_DICARBOXYL\_SYMP.2; 1.  
KW Transport; Transmembrane; Glycoprotein; Symport; Multigene family.  
FT DOMAIN 1 44  
FT TRANSMEM 45 64  
FT TRANSMEM 88 108  
FT TRANSMEM 121 142  
FT TRANSMEM 143 239  
FT DOMAIN 240 259  
FT TRANSMEM 280 301  
FT TRANSMEM 317 339  
FT TRANSMEM 406 430  
FT TRANSMEM 437 459  
FT CARBOHYD 206 206  
FT CONFLICT 216 216  
FT CONFLICT 19 19  
FT CONFLICT 27 27  
FT CONFLICT 50 50  
FT CONFLICT 58 58  
FT CONFLICT 141 141  
FT CONFLICT 155 155  
FT CONFLICT 211 211  
FT CONFLICT 253 255  
FT CONFLICT 263 289  
FT CONFLICT 347 347  
FT CONFLICT 339 339  
FT CONFLICT 556 556  
FT CONFLICT 563 565  
FT CONFLICT 570 570  
SQ SEQUENCE 574 AA; 62104 MW; C8104B6727979435 CRC64;  
PERDQNEVSDHDDHVL (IN REF. 3).  
F -> L (IN REF. 1).  
Y -> F (IN REF. 3).  
A -> G (IN REF. 3).  
CSV -> RVL (IN REF. 3).  
W -> G (IN REF. 3).  
GIA -> AIP (IN REF. 3).  
AKLMVDFNILEIVMKLVIMIMWYSP -> GQADGGFLQH  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
H -> P (IN REF. 1).  
E -> G (IN REF. 3).  
T -> Q (IN REF. 2).  
A -> S (IN REF. 2).  
A -> G (IN REF. 3).  
P -> A (IN REF. 3).  
V -> E (IN REF. 1).  
PERDQNEVSDHDDHVL (IN REF. 3).

Query Match 78.4%; Score 29; DB 1; Length 574;  
Best Local Similarity 75.0%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
DB 403 LYEAVAAI 410

Search completed: July 18, 2003, 09:57:52  
Job time : 2.49289 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:55:28 ; Search time 0.606635 Seconds  
(without alignments)  
388.015 Million cell updates/sec

Title: US-09-857-308-5

Perfect score: 37

Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_A1:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	% Match	Score	Length	ID	Description
1	37	100.0	363	4	US-09-461-697-81	Sequence 81, Appl
2	37	100.0	412	4	US-09-461-697-77	Sequence 77, Appl
3	30	81.1	20	3	US-08-695-301A-30	Sequence 30, Appl
4	30	81.1	20	3	US-08-695-304C-30	Sequence 30, Appl
5	30	81.1	20	4	US-09-589-768-30	Sequence 30, Appl
6	30	81.1	20	4	US-09-594-845-30	Sequence 30, Appl
7	29	78.4	560	2	US-08-948-569A-10	Sequence 10, Appl
8	29	78.4	560	2	US-09-188-469-10	Sequence 10, Appl
9	29	78.4	560	4	US-09-397-238A-10	Sequence 10, Appl
10	29	78.4	564	2	US-08-948-569A-12	Sequence 12, Appl
11	29	78.4	564	2	US-09-188-469-12	Sequence 12, Appl
12	29	78.4	564	4	US-09-397-238A-12	Sequence 12, Appl
13	29	78.4	573	4	US-09-042-709A-19	Sequence 19, Appl
14	29	78.4	574	1	US-08-140-729A-7	Sequence 7, Appl
15	29	78.4	574	1	US-08-546-666-7	Sequence 7, Appl
16	29	78.4	574	2	US-08-916-745-7	Sequence 7, Appl
17	29	78.4	574	2	US-08-948-569A-4	Sequence 4, Appl
18	29	78.4	574	2	US-08-663-808-6	Sequence 6, Appl
19	29	78.4	574	2	US-09-042-929-7	Sequence 7, Appl
20	29	78.4	574	2	US-08-546-661-7	Sequence 7, Appl
21	29	78.4	574	2	US-09-042-960-7	Sequence 7, Appl
22	29	78.4	574	2	US-09-188-469-4	Sequence 4, Appl
23	29	78.4	574	3	US-09-198-650-7	Sequence 7, Appl
24	29	78.4	574	3	US-09-332-740-6	Sequence 6, Appl
25	29	78.4	574	3	US-09-042-913-7	Sequence 7, Appl
26	29	78.4	574	3	US-09-188-496-6	Sequence 6, Appl
27	29	78.4	574	3	US-09-042-937-7	Sequence 7, Appl

28	29	78.4	574	4	US-09-397-238A-4	Sequence 4, Appl
29	29	78.4	574	4	US-09-368-282-6	Sequence 6, Appl
30	29	78.4	574	4	US-09-566-708A-6	Sequence 6, Appl
31	29	78.4	574	4	US-09-042-709A-7	Sequence 7, Appl
32	28	75.7	524	1	US-08-529-654-2	Sequence 2, Appl
33	28	75.7	524	1	US-08-529-654-4	Sequence 4, Appl
34	28	75.7	524	4	US-09-042-709A-20	Sequence 20, Appl
35	28	75.7	525	1	US-08-140-729A-9	Sequence 9, Appl
36	28	75.7	525	1	US-08-546-666-9	Sequence 9, Appl
37	28	75.7	525	2	US-08-916-745-9	Sequence 9, Appl
38	28	75.7	525	2	US-08-948-569A-6	Sequence 6, Appl
39	28	75.7	525	2	US-08-663-808-8	Sequence 8, Appl
40	28	75.7	525	2	US-09-042-929-9	Sequence 9, Appl
41	28	75.7	525	2	US-08-546-661-9	Sequence 9, Appl
42	28	75.7	525	2	US-09-042-960-9	Sequence 9, Appl
43	28	75.7	525	2	US-09-188-469-6	Sequence 6, Appl
44	28	75.7	525	3	US-09-198-650-9	Sequence 9, Appl
45	28	75.7	525	3	US-09-332-740-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-09-461-697-81  
; Sequence 81, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Purnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 81  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-461-697-81

Query Match 100.0%; Score 37; DB 4; Length 363;  
Best Local Similarity 100.0%; Pred. No. 0.89;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
| | | | | | | |  
Db 107 LYQAVATI 114

RESULT 2  
US-09-461-697-77  
; Sequence 77, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Purnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 81  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-461-697-77

FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 77  
LENGTH: 412  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-461-697-77

Query Match 100.0%; Score 37; DB 4; Length 412;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
|||  
Db 156 LYQAVATI 163

RESULT 3  
US-08-695-301A-30  
Sequence 30, Application US/08695301A  
Patent No. 6093400  
GENERAL INFORMATION:  
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN  
TITLE OF INVENTION: Modified HGP-30 Peptides, Conjugates,  
TITLE OF INVENTION: Compositions and Methods of Use  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Office of Sherman and Shalloway  
STREET: 413 N. Washington Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage  
MEDIUM TYPE: 3.5 inch, 1.44 mb storage  
COMPUTER: Dell System 210; Intel 80 286 Microprocessor  
OPERATING SYSTEM: MS DOS 6.22  
SOFTWARE: Word Perfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/695,301A  
FILING DATE: August 9, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Richard A. Steinberg  
REGISTRATION NUMBER: 26,588  
REFERENCE/DOCKET NUMBER: CELL-101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 549-2282  
TELEFAX: (703) 836-0106  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal fragment  
FEATURE:

NAME/KEY:  
LOCATION: 75 to 94  
OTHER INFORMATION: fragment of p-17 gag protein of  
HIV-1RF  
US-08-695-301A-30

Query Match 81.1%; Score 30; DB 3; Length 20;  
Best Local Similarity 75.0%; Pred. No. 1.3;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
|||  
Db 3 LYNAVATL 10

RESULT 4  
US-08-695-304C-30  
Sequence 30, Application US/08695304C  
Patent No. 6103239  
GENERAL INFORMATION:  
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN  
TITLE OF INVENTION: Modified HGP-30 Heteroconjugates,  
TITLE OF INVENTION: Compositions and Methods of Use  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Office of Sherman and Shalloway  
STREET: 413 N. Washington Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage  
MEDIUM TYPE: 3.5 inch, 1.44 mb storage  
COMPUTER: Dell System 210; Intel 80 286 Microprocessor  
OPERATING SYSTEM: MS DOS 6.22  
SOFTWARE: Word Perfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/695,304C  
FILING DATE: August 9, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Richard A. Steinberg  
REGISTRATION NUMBER: 26,588  
REFERENCE/DOCKET NUMBER: CELL-102  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 549-2282  
TELEFAX: (703) 836-0106  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal fragment  
FEATURE:  
NAME/KEY:  
LOCATION: 75 to 94  
OTHER INFORMATION: fragment of p-17 gag protein of  
HIV-1RF  
US-08-695-304C-30

Query Match 81.1%; Score 30; DB 3; Length 20;  
Best Local Similarity 75.0%; Pred. No. 1.3;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
|||  
Db 3 LYNAVATL 10

RESULT 5  
US-09-589-768-30  
Sequence 30, Application US/09589768  
Patent No. 6268472  
GENERAL INFORMATION:  
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN  
TITLE OF INVENTION: Modified HGP-30 Peptides, Conjugates,  
TITLE OF INVENTION: Compositions and Methods of Use  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Office of Sherman and Shalloway  
STREET: 413 N. Washington Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage  
MEDIUM TYPE: 3.5 inch, 1.44 mb storage  
COMPUTER: Dell System 210; Intel 80 286 Microprocessor  
OPERATING SYSTEM: MS DOS 6.22  
SOFTWARE: Word Perfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/589,768  
FILING DATE: Concurrently herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Richard A. Steinberg  
REGISTRATION NUMBER: 26,588  
REFERENCE/DOCKET NUMBER: CELL-101.01  
TELEPHONE: (703) 549-2282  
TELEFAX: (703) 836-0106  
INFORMATION FOR SEQ ID NO: 30:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal fragment  
FEATURE:  
NAME/KEY: 75 to 94  
LOCATION: fragment of p-17 gag protein of  
OTHER INFORMATION: HIV-1RF  
US-09-589-768-30

Query Match 81.1%; Score 30; DB 4; Length 20;  
Best Local Similarity 75.0%; Pred. No. 1.3;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
|||  
Db 3 LYNAAVATL 10

RESULT 6  
US-09-594-845-30  
Sequence 30, Application US/09594845  
Patent No. 6287565  
GENERAL INFORMATION:  
APPLICANT: DANIEL H. ZIMMERMAN, PREM S. SARIN  
TITLE OF INVENTION: Modified HGP-30 Heteroconjugates,  
TITLE OF INVENTION: Compositions and Methods of Use  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Office of Sherman and Shalloway  
STREET: 413 N. Washington Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 1.2 mb storage  
MEDIUM TYPE: 3.5 inch, 1.44 mb storage  
COMPUTER: Dell System 210; Intel 80 286 Microprocessor  
OPERATING SYSTEM: MS DOS 6.22  
SOFTWARE: Word Perfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/594,845  
FILING DATE: Concurrently herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Richard A. Steinberg  
REGISTRATION NUMBER: 26,588  
REFERENCE/DOCKET NUMBER: CELL-102.01  
TELEPHONE: (703) 549-2282  
TELEFAX: (703) 836-0106  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal fragment  
FEATURE:  
NAME/KEY: 75 to 94  
LOCATION: fragment of p-17 gag protein of  
OTHER INFORMATION: HIV-1RF  
US-09-594-845-30

Query Match 81.1%; Score 30; DB 4; Length 20;  
Best Local Similarity 75.0%; Pred. No. 1.3;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
|||  
Db 3 LYNAAVATL 10

RESULT 7  
US-08-948-569A-10  
Sequence 10, Application US/08948569A  
Patent No. 5882926  
GENERAL INFORMATION:  
APPLICANT: Amara, Susan G  
APPLICANT: Arriza, Jeffrey L  
APPLICANT: Eliasof, Scott  
APPLICANT: Kavanaugh, Michael P  
TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes  
TITLE OF INVENTION: and Uses  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDonnell-Boehnen Hulbert & Berghoff  
STREET: 300 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,569A  
FILING DATE: 10-OCT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5882926nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,509-F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 560 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-948-569A-10

Query Match 78.4%; Score 29; DB 2; Length 560;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
|||  
Db 383 LYEAAAI 390

## RESULT 8

```

US-09-188-469-10
; Sequence 10, Application US/09188469
; Patent No. 5989825
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
;

```

```
Query Match      78.4%; Score 29; DB 2; Length 560;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 1; Indels
```

Qy 1 LYQAVATI 8  
||:||||  
Db 383 LYEAVAAI 390

## RESULT 9

US-09-397-238A-10  
 ; Sequence 10, Application US/09397238A  
 ; Patent No. 6284505  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: Amara, Susan G  
 ; Arriza, Jeffrey L  
 ; Eliasof, Scott  
 ; Kavanaugh, Michael P  
 ;  
 ; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes  
 ; and Uses  
 ;  
 ; NUMBER OF SEQUENCES: 18  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: McDonnell Boehnen Hulbert & Berghoff  
 ; STREET: 300 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA

```
,  
,  
,  
ZIP: 60606  
  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
  
CURRENT APPLICATION DATA: US/09/397,238A  
FILING DATE: 16-Sep-1999  
CLASSIFICATION: <Unknown>  
  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6284505nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,509-F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002
```

Query Match 78.4%; Score 29; DB 4; Length 560;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 1; Indels

QY 1 LYQAVATI 8  
||:|||||  
Db 383 LYEAVAAI 390

RESULT 10

US-08-948-569A-12  
 Sequence 12, Application US/08948569A  
 Patent No. 5882926  
 GENERAL INFORMATION:  
 APPLICANT: Amara, Susan G  
 APPLICANT: Arriza, Jeffrey L  
 APPLICANT: Eliasof, Scott  
 APPLICANT: Kavanaugh, Michael P  
 TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes  
 TITLE OF INVENTION: and Uses  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
 STREET: 300 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60606

```

? ZIP: 08000
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/948,569A
? FILING DATE: 10-OCT-1997
? CLASSIFICATION: 435
?

```

:		TITLE OF INVENTION
:	KAVAH	
:		NUMBER OF SEQUENCES
:		

;  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-948-569A-12

Query Match 78.4%; Score 29; DB 2; Length 564;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
||:|||||  
Db 385 LYEAVAAI 392

## RESULT 11

US-09-188-469-12  
; Sequence 12, Application US/09188469  
; Patent No. 5989825

;  
; GENERAL INFORMATION:  
; APPLICANT: Amara, Susan G  
; APPLICANT: Arriza, Jeffrey L  
; APPLICANT: Eliasof, Scott  
; APPLICANT: Kavanaugh, Michael P  
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes  
; TITLE OF INVENTION: and Uses  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606

;  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/188,469  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/948,569  
; FILING DATE: 10-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5989825nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 93,509-F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 12;  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 564 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-188-469-12

Query Match 78.4%; Score 29; DB 2; Length 564;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
||:|||||  
Db 385 LYEAVAAI 392

## RESULT 12

US-09-397-238A-12  
; Sequence 12, Application US/09397238A  
; Patent No. 6284505

;  
; GENERAL INFORMATION:  
; APPLICANT: Amara, Susan G  
; APPLICANT: Arriza, Jeffrey L  
; APPLICANT: Eliasof, Scott  
; APPLICANT: Kavanaugh, Michael P  
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes  
; TITLE OF INVENTION: and Uses  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606

;  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/397,238A  
; FILING DATE: 16-Sep-1999  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6284505nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 93,509-F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 12;  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 564 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-09-397-238A-12

Query Match 78.4%; Score 29; DB 4; Length 564;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
||:|||||  
Db 385 LYEAVAAI 392

## RESULT 13

US-09-042-709A-19  
; Sequence 19, Application US/09042709A  
; Patent No. 6458571  
; GENERAL INFORMATION:  
; APPLICANT: State of Oregon  
; APPLICANT: Amara, Susan G  
; APPLICANT: Arriza, Jeffrey L  
; TITLE OF INVENTION: AMINO ACID TRANSPORTERS AND USES  
; FILE REFERENCE: 93,509-H (99/145)  
; CURRENT APPLICATION NUMBER: US/09/042,709A  
; CURRENT FILING DATE: 1998-03-17  
; PRIOR APPLICATION NUMBER: 140,729  
; PRIOR FILING DATE: 1993-10-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 19  
; LENGTH: 573  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-042-709A-19

Query Match 78.4%; Score 29; DB 4; Length 573;  
Best Local Similarity 75.0%; Pred. No. 1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
||:|||||  
Db 402 LYEAVAAI 409

## RESULT 14

US-08-140-729A-7  
; Sequence 7, Application US/08140729A  
; Patent No. 5658782  
; GENERAL INFORMATION:  
; APPLICANT: Amara, Susan G  
; APPLICANT: Arriza, Jeffrey L  
; TITLE OF INVENTION: Amino Acid Transporters and Uses  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/140,729A  
; FILING DATE: 20 OCT 1993  
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5658782nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 93,509  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 574 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-140-729A-7

Query Match 78.4%; Score 29; DB 1; Length 574;  
Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
||:|||||  
Db 403 LYEAVAAI 410

## RESULT 15

US-08-546-666-7  
; Sequence 7, Application US/08546666  
; Patent No. 5776774  
; GENERAL INFORMATION:  
; APPLICANT: Amara, Susan G  
; APPLICANT: Arriza, Jeffrey L  
; TITLE OF INVENTION: Amino Acid Transporters and Uses  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/546,666  
; FILING DATE:  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/140,729  
; FILING DATE: 20 OCT 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5776774nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 93,509  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 574 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-546-666-7

Query Match 78.4%; Score 29; DB 1; Length 574;  
Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
||:|||||  
Db 403 LYEAVAAI 410

Search completed: July 18, 2003, 10:01:23  
Job time : 1.60664 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:57:18 ; Search time 1.1564 Seconds  
(without alignments)  
821.584 Million cell updates/sec

Title: US-09-857-308-5  
Perfect score: 37  
Sequence: 1 LYQAVATI 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap1:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap2:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap3:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	37	100.0	363	11	US-09-922-261-81	Sequence 81, Appl
2	37	100.0	412	11	US-09-922-261-77	Sequence 77, Appl
3	32	96.5	897	15	US-10-043-487-287	Sequence 287, Appl
4	35	86.5	273	15	US-10-027-806-56	Sequence 56, Appl
5	32	86.5	273	15	US-10-034-623-56	Sequence 56, Appl
6	32	86.5	273	15	US-10-027-801-56	Sequence 56, Appl
7	31	83.8	269	15	US-10-027-806-24	Sequence 24, Appl
8	31	83.8	269	15	US-10-034-623-24	Sequence 24, Appl
9	31	83.8	269	15	US-10-027-801-24	Sequence 24, Appl
10	29	78.4	481	11	US-09-815-923-8	Sequence 8, Appl
11	29	78.4	574	10	US-09-748-657-2	Sequence 2, Appl
12	29	78.4	574	12	US-09-040-736-2	Sequence 2, Appl
13	29	78.4	574	15	US-10-261-399-7	Sequence 7, Appl
14	29	78.4	760	12	US-09-759-1308-440	Sequence 440, App
15	29	78.4	760	12	US-09-759-1308-446	Sequence 446, App
16	29	78.4	760	15	US-10-042-431-70	Sequence 70, Appl

17	29	78.4	760	15	US-10-042-431-76	Sequence 76, Appl
18	28	75.7	223	15	US-10-156-761-9526	Sequence 9526, Ap
19	28	75.7	330	15	US-10-156-761-13865	Sequence 13865, A
20	28	75.7	333	15	US-10-156-761-7860	Sequence 7860, Ap
21	28	75.7	335	12	US-09-940-244-349	Sequence 349, App
22	28	75.7	412	15	US-10-101-464A-905	Sequence 905, App
23	28	75.7	482	15	US-10-101-464A-955	Sequence 955, App
24	28	75.7	525	15	US-10-261-399-9	Sequence 9, Appli
25	28	75.7	712	10	US-09-925-302-678	Sequence 678, App
26	28	75.7	997	15	US-10-101-464A-977	Sequence 977, App
27	27	73.0	137	15	US-10-156-761-12625	Sequence 12625, A
28	27	73.0	183	11	US-09-738-626-6050	Sequence 6050, Ap
29	27	73.0	209	10	US-09-764-870-323	Sequence 323, App
30	27	73.0	209	10	US-09-764-853-546	Sequence 546, App
31	27	73.0	209	15	US-10-125-540-323	Sequence 323, App
32	27	73.0	209	15	US-10-103-313-364	Sequence 364, App
33	27	73.0	339	12	US-09-791-489-2	Sequence 2, Appli
34	27	73.0	380	12	US-09-782-816A-56	Sequence 56, Appl
35	27	73.0	535	10	US-09-815-242-13419	Sequence 2, Appli
36	27	73.0	535	11	US-09-844-199-2	Sequence 4, Appli
37	27	73.0	611	15	US-10-074-547-4	Sequence 1, Appli
38	27	73.0	2502	10	US-09-772-316-1	Sequence 3, Appli
39	26	70.3	9	8	US-08-600-483-3	Sequence 3, Appli
40	26	70.3	9	10	US-09-789-720-3	Sequence 85, Appl
41	26	70.3	9	11	US-09-912-787-85	Sequence 2, Appli
42	26	70.3	9	11	US-09-954-392-2	Sequence 80, Appl
43	26	70.3	9	15	US-10-014-326-80	Sequence 25, Appl
44	26	70.3	9	15	US-10-106-487-25	Sequence 21, Appl
45	26	70.3	9	15	US-10-114-823B-21	

ALIGNMENTS

RESULT 1  
US-09-922-261-81  
; Sequence 81, Application US/09922261  
; Patent No. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US/09/461,697  
; PRIOR FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 81  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-261-81

Query Match 100.0%; Score 37; DB 11; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LYQAVATI 8  
Db 107 LYQAVATI 114

RESULT 2  
US-09-922-261-77

; Sequence 77, Application US/09922261  
; Patent No. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Purnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US/09/461,697  
; PRIOR FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 77  
; LENGTH: 412  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-261-77

Query Match 100.0%; Score 37; DB 11; Length 412;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 156 LYQAVATI 163

## RESULT 3

US-10-043-487-287  
; Sequence 287, Application US/10043487  
; Publication No. US2003005220A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, LERAIN  
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptide  
; TITLE OF INVENTION: mammalian polypeptides  
; FILE REFERENCE: B4778A  
; CURRENT APPLICATION NUMBER: US/10/043,487  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/261,130  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 561  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 287  
; LENGTH: 897  
; TYPE: PRT  
; ORGANISM: Shigella Flexneri  
US-10-043-487-287

Query Match 94.6%; Score 35; DB 15; Length 897;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 770 LYQAVATI 777

## RESULT 4

US-10-027-806-56  
; Sequence 56, Application US/10027806  
; Publication No. US20020160476A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,801  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0

; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,806  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 56  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-027-806-56

Query Match 86.5%; Score 32; DB 15; Length 273;  
Best Local Similarity 87.5%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 178 LYQAVATI 185

## RESULT 5

US-10-034-623-56  
; Sequence 56, Application US/10034623  
; Publication No. US20020198365A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/034,623  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 60/102,294  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 56  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-034-623-56

Query Match 86.5%; Score 32; DB 15; Length 273;  
Best Local Similarity 87.5%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 178 LYQAVATI 185

## RESULT 6

US-10-027-801-56  
; Sequence 56, Application US/10027801  
; Publication No. US20030054364A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,801  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 56  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-027-801-56

Query Match 86.5%; Score 32; DB 15; Length 273;  
Best Local Similarity 87.5%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 178 LYQAVPTI 185  
|||||

RESULT 7  
US-10-027-806-24  
; Sequence 24, Application US/10027806  
; Publication No. US20020160476A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCOIP.002A  
; CURRENT APPLICATION NUMBER: US/10/027.806  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 24  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-027-806-24

Query Match 83.8%; Score 31; DB 15; Length 269;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 174 LYQAVPTV 181  
|||||

RESULT 8  
US-10-034-623-24  
; Sequence 24, Application US/10034623  
; Publication No. US20020198365A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCOIP.002A  
; CURRENT APPLICATION NUMBER: US/10/034.623  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 60/102,294  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 24  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-034-623-24

Query Match 83.8%; Score 31; DB 15; Length 269;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 174 LYQAVPTV 181  
|||||

RESULT 9  
US-10-027-801-24  
; Sequence 24, Application US/10027801  
; Publication No. US20030054364A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCOIP.002A  
; CURRENT APPLICATION NUMBER: US/10/027.801  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 24  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-027-801-24

Query Match 83.8%; Score 31; DB 15; Length 269;  
Best Local Similarity 75.0%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 174 LYQAVPTV 181  
|||||

RESULT 10  
US-09-815-923-8  
; Sequence 8, Application US/09815923  
; Publication No. US20020197644A1  
; GENERAL INFORMATION:  
; APPLICANT: Gill, Sarjeet S.  
; APPLICANT: Ross, Linda S.  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US20020197644A  
; FILE REFERENCE: Target Sites for Insecticides  
; CURRENT APPLICATION NUMBER: US/09/815.923  
; CURRENT FILING DATE: 2001-03-23  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Aedes aegypti  
; FEATURE:  
; OTHER INFORMATION: glutamate transporter  
US-09-815-923-8

Query Match 78.4%; Score 29; DB 11; Length 481;  
Best Local Similarity 75.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LYQAVATI 8  
Db 353 LYEAVAAI 360  
|||||

RESULT 11  
US-09-748-657-2  
; Sequence 2, Application US/09748657  
; Patent No. US20020099025A1

GENERAL INFORMATION:  
; APPLICANT: HEYWOOD, JAMES  
; TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS  
; FILE REFERENCE: 102243-8  
; CURRENT APPLICATION NUMBER: US/09/748,657  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,764  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 574  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-748-657-2

Query Match 78.4%; Score 29; DB 10; Length 574;  
Best Local Similarity 75.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LYQAVATI 8  
Db 403 LYEAVAAI 410

RESULT 12  
US-09-040-736-2  
; Sequence 2, Application US/09040736  
; Publication No. US20030060617A1  
GENERAL INFORMATION:  
; APPLICANT: Rothstein, Jeffery D.  
; APPLICANT: Lin, Chieng-Liang Glenn  
; APPLICANT: Bristol, Lynn A.  
; TITLE OF INVENTION: METHODS FOR DETECTING NEUROLOGICAL  
; TITLE OF INVENTION: DISORDERS  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/040.736  
; FILING DATE: 18-MAR-1998

ATTORNEY/AGENT INFORMATION:  
; NAME: Corless, Peter F.  
; REGISTRATION NUMBER: 33,860  
; REFERENCE/DOCKET NUMBER: 1699/48148  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400  
; TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 574 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-040-736-2

Query Match 78.4%; Score 29; DB 12; Length 574;  
Best Local Similarity 75.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LYQAVATI 8  
Db 403 LYEAVAAI 410

RESULT 13  
US-10-261-399-7  
; Sequence 7, Application US/10261399  
; Publication No. US20030125538A1  
GENERAL INFORMATION:  
; APPLICANT: Amara, Susan G  
; Arriza, Jeffery L  
; TITLE OF INVENTION: Amino Acid Transporters and Uses  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allegretti & Witcoff, Ltd.  
; STREET: 10 South Wacker Drive, Suite 3000  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/261,399  
; FILING DATE: 01-Oct-2002  
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/140,729  
; FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:  
; NAME: NO. US20030125538A1nan, Kevin E  
; REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 574 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-10-261-399-7

Query Match 78.4%; Score 29; DB 15; Length 574;  
Best Local Similarity 75.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 LYQAVATI 8  
Db 403 LYEAVAAI 410

RESULT 14  
US-09-759-130B-440  
; Sequence 440, Application US/09759130B  
; Publication No. US20030022279A1  
GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: McCarthy, Sean A  
; APPLICANT: Fraser, Christopher C  
; APPLICANT: Sharp, John D  
; APPLICANT: Barnes, Thomas S  
; APPLICANT: Kirst, Susan J  
; APPLICANT: Mackay, Charles R  
; APPLICANT: Myers, Paul S  
; APPLICANT: Leiby, Kevin R  
; APPLICANT: Wrighton, Nicolas  
; APPLICANT: Goodearl, Andrew  
; APPLICANT: Holtzman, Douglas A

;; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING  
;; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER  
;; FILE REFERENCE: MPI00-5350NMIM  
;; CURRENT APPLICATION NUMBER: US/09/759,130B  
;; CURRENT FILING DATE: 2002-09-16  
;; PRIOR APPLICATION NUMBER: US 09/479,249  
;; PRIOR FILING DATE: 2000-01-07  
;; PRIOR APPLICATION NUMBER: US 09/559,497  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: US 09/578,063  
;; PRIOR FILING DATE: 2000-05-24  
;; PRIOR APPLICATION NUMBER: US 09/333,159  
;; PRIOR FILING DATE: 1999-06-14  
;; PRIOR APPLICATION NUMBER: US 09/596,194  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 09/342,364  
;; PRIOR FILING DATE: 1999-06-29  
;; PRIOR APPLICATION NUMBER: US 09/608,452  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/393,996  
;; PRIOR FILING DATE: 1999-09-10  
;; PRIOR APPLICATION NUMBER: US 09/602,871  
;; PRIOR FILING DATE: 2000-06-23  
;; PRIOR APPLICATION NUMBER: US 09/420,707  
;; NUMBER OF SEQ ID NOS: 460  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 440  
;; LENGTH: 760  
;; TYPE: PRT  
;; ORGANISM: Mus sp.  
US-09-759-130B-440

Query Match 78.4%; Score 29; DB 12; Length 760;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 LYQAVAT 7  
Db 620 LYQCVAT 626  
||| |||

RESULT 15  
US-09-759-130B-446  
;; Sequence 446, Application US/09/59130B  
;; Publication No. US20030022279A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Millennium Pharmaceuticals, Inc.  
;; APPLICANT: McCarthy, Sean A  
;; APPLICANT: Fraser, Christopher C  
;; APPLICANT: Sharp, John D  
;; APPLICANT: Barnes, Thomas S  
;; APPLICANT: Kirt, Susan J  
;; APPLICANT: Mackay, Charles R  
;; APPLICANT: Myers, Paul S  
;; APPLICANT: Leiby, Kevin R  
;; APPLICANT: Wrighton, Nicolas  
;; APPLICANT: Goodearl, Andrew  
;; APPLICANT: Holtzman, Douglas A  
;; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING  
;; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER  
;; FILE REFERENCE: MPI00-5350NMIM  
;; CURRENT APPLICATION NUMBER: US/09/759,130B  
;; CURRENT FILING DATE: 2002-09-16  
;; PRIOR APPLICATION NUMBER: US 09/479,249  
;; PRIOR FILING DATE: 2000-01-07  
;; PRIOR APPLICATION NUMBER: US 09/559,497  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: US 09/578,063  
;; PRIOR FILING DATE: 2000-05-24  
;; PRIOR APPLICATION NUMBER: US 09/333,159

;; PRIOR FILING DATE: 1999-06-14  
;; PRIOR APPLICATION NUMBER: US 09/596,194  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 09/342,364  
;; PRIOR FILING DATE: 1999-06-29  
;; PRIOR APPLICATION NUMBER: US 09/608,452  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/393,996  
;; PRIOR FILING DATE: 1999-09-10  
;; PRIOR APPLICATION NUMBER: US 09/602,871  
;; PRIOR FILING DATE: 2000-06-23  
;; PRIOR APPLICATION NUMBER: US 09/420,707  
;; NUMBER OF SEQ ID NOS: 460  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 446  
;; LENGTH: 760  
;; TYPE: PRT  
;; ORGANISM: Mus sp.  
US-09-759-130B-446  
Query Match 78.4%; Score 29; DB 12; Length 760;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 LYQAVAT 7  
Db 620 LYQCVAT 626  
||| |||

Search completed: July 18, 2003, 10:02:40  
Job time : 3.1564 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:41:22 ; Search time 399 Seconds  
(without alignments)  
8846.602 Million cell updates/sec

Title: US-09-857-308-2  
Perfect score: 1711  
Sequence: 1 acgcgactctgctcaggc.....aaaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues  
Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq1.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq3.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	75.3	1825	11	US-09-922-261-75
2	1233.2	72.1	1239	11	US-09-922-261-76
3	1092	63.8	1092	11	US-09-922-261-80
4	600	35.1	600	11	US-09-922-261-92
5	510	29.8	510	11	US-09-922-261-96
6	505.8	29.6	573	12	US-09-918-995-14733
7	374.2	21.9	379	12	US-09-918-995-36873
8	337	19.7	491	12	US-09-918-995-31390
9	336	19.6	336	11	US-09-922-261-100
10	293.8	17.2	517	11	US-09-764-847-352
11	293.8	17.2	517	15	US-10-092-154-352
12	230.6	13.5	414	12	US-09-918-995-3654
13	216	12.6	216	11	US-09-922-261-104
14	159	9.3	159	11	US-09-922-261-94
15	141	8.2	141	11	US-09-922-261-112
16	132	7.7	132	11	US-09-922-261-114

17	107	6.3	174	11	US-09-783-590-3427	Sequence 3427, Ap
18	105	6.1	105	11	US-09-922-261-78	Sequence 78, Appl
19	96	5.6	96	11	US-09-922-261-88	Sequence 88, Appl
20	90	5.3	90	11	US-09-922-261-116	Sequence 116, Appl
21	87	5.1	87	11	US-09-922-261-90	Sequence 90, Appl
22	72	4.2	72	11	US-09-922-261-118	Sequence 118, Appl
23	69	4.0	69	11	US-09-922-261-84	Sequence 84, Appl
24	60	3.5	60	11	US-09-922-261-106	Sequence 106, Appl
25	57	3.3	57	11	US-09-922-261-120	Sequence 120, Appl
26	51	3.0	51	11	US-09-922-261-98	Sequence 98, Appl
27	48	2.8	48	11	US-09-922-261-108	Sequence 108, Appl
28	44	2.6	1654	10	US-09-925-302-156	Sequence 156, Appl
29	42.6	2.5	712	15	US-10-198-846-8022	Sequence 8022, Ap
30	42.2	2.5	1708	15	US-10-106-698-2124	Sequence 2124, Ap
c 31	42	2.5	520	11	US-09-834-975-635	Sequence 635, Appl
32	42	2.5	2329	10	US-09-800-729-11	Sequence 11, Appl
33	42	2.5	2355	10	US-09-800-729-27	Sequence 27, Appl
34	42	2.5	2500	15	US-10-102-806-88	Sequence 88, Appl
35	41.8	2.4	903	15	US-10-198-846-4296	Sequence 4296, Ap
c 36	41.4	2.4	168	11	US-09-964-824A-134	Sequence 134, Appl
c 37	41.2	2.4	308	11	US-09-960-352-7670	Sequence 7670, Ap
38	41	2.4	1195	15	US-10-180-375-25	Sequence 25, Appl
39	41	2.4	2823	15	US-10-103-313-50	Sequence 50, Appl
c 40	40.8	2.4	892	15	US-10-198-846-7154	Sequence 7154, Ap
c 41	40.4	2.4	461	15	US-10-198-846-6011	Sequence 6011, Ap
42	40.2	2.3	549	12	US-09-991-936-1736	Sequence 1736, Ap
c 43	40	2.3	355	15	US-10-198-846-10174	Sequence 10174, A
44	39.6	2.3	312	12	US-09-907-907A-14	Sequence 14, Appl
45	39.6	2.3	439	12	US-09-918-995-13988	Sequence 13988, A

## ALIGNMENTS

RESULT 1  
US-09-922-261-75  
; Sequence 75, Application US/09922261  
; Patent No. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 75  
; LENGTH: 1825  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-922-261-75

Query Match	75.3%	Score	1288;	DB	11;	Length	1825;
Best Local Similarity	99.6%	Pred. No.	0;				
Matches	1291;	Conservative	0;	Mismatches	5;	Indels	0;
Gaps	0;						
Qy	344	AATCTGCAAGACTCTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC	403				
Db	26	AAATGTTGAGACTCTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC	85				
Qy	404	TTGATTTGCTCCACCGGAGTTCCGTCGTGGTGAAGTCCATGACCCACCCCTGCACCAA	463				
Db	86	TTGATTTGCTCCACCGGAGTTCCGTCGTGGTGAAGTCCATGACCCACCCCTGCACCAA	145				

```
QY 464 CCTCAGCCACAAGCCGAAGCCGCCACTATGCTGGACATCCCTCAGAGCCATGTAGT 523
Db |||||||
QY 146 CCTCAGCCACAAGCCGAAGCCGCCACTATGCTGGACATCCCTCAGAGCCATGTAGT 205
Db |||||||
QY 524 CTCACCATCATACGATTGATTCAGCACAACCGACGCTCTCGAACCTTATGCC 583
Db |||||||
QY 206 CTCACCATCATACGATTGATTCAGCACAACCGACGCTCTCGAACCTTATGCC 265
Db |||||||
QY 584 ACAGCTCAGGCCCCAGAAATCAGCAGCAGACAGAGAGGTGTAAAACTGAAGAGTGAACCT 643
Db |||||||
QY 266 ACAGCTCAGGCCCCAGAAATCAGCAGCAGACAGAGAGGTGTAAAACTGAAGAGTGAACCT 325
Db |||||||
QY 644 CTTCCCTCGTGGCCCTGAGTACCTCTCTCCCTGATGACCTCTCGCTTTAGATTGAAG 703
Db |||||||
QY 326 CTTCCCTCGTGGCCCTGAGTACCTCTCTCCCTGATGACCTCTCGCTTTAGATTGAAG 385
Db |||||||
QY 704 AATCCCAATGCACCATTCAGATCCGGCACAGTGCACAGTGCACAGAGTGAATTTATCGTGG 763
Db |||||||
QY 386 AATCCCAATGCACCATTCAGATCCGGCACAGTGCACAGTGCACAGAGTGAATTTATCGTGG 445
Db |||||||
QY 764 AAAGGGAACTGTGACTGAACTCAGCTGGCACTCTCTGCGCAGCTCTCTACCAAGCA 823
Db |||||||
QY 446 AAAGGGAACTGTGACTGAACTCAGCTGGCACTCTCTGCGCAGCTCTCTACCAAGCA 505
Db |||||||
QY 824 GTGGCCAAATCCTTGCCCCACCGGGCTTTGACTGTGCTTAATGAGAGTGTCTGGAGACC 883
Db |||||||
QY 506 GTGGCCAAATCCTTGCCCCACCGGGCTTTGACTGTGCTTAATGAGAGTGTCTGGAGACC 565
Db |||||||
QY 884 CTAAGTGTGGCAGATGATTTGCCCTTAAGTTTACCAAGTGTCTGGTGTGCTGTG 943
Db |||||||
QY 566 CTAAGTGTGGCAGATGATTTGCCCTTAAGTTTACCAAGTGTCTGGTGTGCTGTG 625
Db |||||||
QY 944 GACCGGGAGCCGCGCTGGCAGACAGTCTCTTTCTGATGTGATGGAGCAGTATTCCAT 1003
Db |||||||
QY 626 GACCGGGAGCCGCGCTGGCAGACAGTCTCTTTCTGATGTGATGGAGCAGTATTCCAT 585
Db |||||||
QY 1004 GAAGTGGGTATTGGCAGTGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCAGCATCAAGAC 1063
Db |||||||
QY 686 GAAGTGGGTATTGGCAGTGTGCTCTCCCTCCAGAAAGTTCTGGCAGCACCAGCATCAAGAC 745
Db |||||||
QY 1064 TATCAGATTACATCTACAGATTAGTAGCAACTCTCTGAGAAATATGAAGGATTGTC 1123
Db |||||||
QY 746 TATCAGATTACATCTACAGATTAGTAGCAACTCTCTGAGAAATATGAAGGATTGTC 805
Db |||||||
QY 1124 AATCTGTAGAAGCCACAGAGGACGCTAAACCTGTGAAGATCAAGAGGAACTGTGAGC 1183
Db |||||||
QY 806 AATCTGTAGAAGCCACAGAGGACGCTAAACCTGTGAAGATCAAGAGGAACTGTGAGC 865
Db |||||||
QY 1184 GACATCACTTTTCTGTAGTGAGGAGCTGGAGGCTGACCTTGTCTTGAGACCAAGTCA 1243
Db |||||||
QY 866 GACATCACTTTTCTGTAGTGAGGAGCTGGAGGCTGACCTTGTCTTGAGACCAAGTCA 925
Db |||||||
QY 1244 CTGCTATGGGAGTGTGGGGCTCAGACGCAACCTTCCCTCAACTCAAGCTGAGGTTGAA 1303
Db |||||||
QY 926 CTGCTATGGGAGTGTGGGGCTCAGACGCAACCTTCCCTCAACTCAAGCTGAGGTTGAA 985
Db |||||||
QY 1304 GCTTCAACACAGGCTTCAAGTGCAGAGGTAAATGCTTCTCTTTTGAATCTGGCCCAT 1363
Db |||||||
QY 986 GCTTCAACACAGGCTTCAAGTGCAGAGGTAAATGCTTCTCTTTTGAATCTGGCCCAT 1045
Db |||||||
QY 1364 GTGAAATGGAGCCCTCAGAAAGTGAAGAGCAATGCTCTGGCATGGTGTGGGCT 1423
Db |||||||
QY 1046 GTGAAATGGAGCCCTCAGAAAGTGAAGAGCAATGCTCTGGCATGGTGTGGGCT 1105
Db |||||||
QY 1424 AGTGATGTCTTCAGAGGCTATGTCAGGATGAGTGAAGCTGGATTCCTCAGAGCCCT 1483
Db |||||||
QY 1106 AGTGATGTCTTCAGAGGCTATGTCAGGATGAGTGAAGCTGGATTCCTCAGAGCCCT 1165
Db |||||||
QY 1484 GATGACTCAGATGAGGCTATGTTCCCACTCCACTGACAGGCTCATGGGTCCTCCCT 1543
Db |||||||
QY 1166 GATGACTCAGATGAGGCTATGTTCCCACTCCACTGACAGGCTCATGGGTCCTCCCT 1225
Db |||||||
```

## RESULT 2

```
US-09-922-261-76
; Sequence 76, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-76
```

```
Query Match 72.1%; Score 1233.2; DB 11; Length 1239;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 348 TCACAAGTACTGGGGAGAGATACCAATATCATCAAGCCAGACCAACAGACTTCTCTCG 407
Db |||||||
QY 2 TGTTCAGATACATGGGGAGAGATACCAATATCATCAAGCCAGACCAACAGACTTCTCTCG 61
Db |||||||
QY 408 ATTTGCTCCACGGGAGTTCCGTCGTGGTGAAGTCCATGACACCCCTGCACCAACCT 467
Db |||||||
QY 62 ATTTGCTCCACGGGAGTTCCGTCGTGGTGAAGTCCATGACACCCCTGCACCAACCT 121
Db |||||||
QY 468 CAGCCACAAGCCGAAGCCGCCACTATGCTGGACATCCCTCAGAGCATGTAGTCTCA 527
Db |||||||
QY 122 CAGCCACAAGCCGAAGCCGCCACTATGCTGGACATCCCTCAGAGCATGTAGTCTCA 181
Db |||||||
QY 528 CCATCCATACGATTGATTCAGCACAACCCAGCTCTTCGCAACCTTATTGCCACAG 587
Db |||||||
QY 182 CCATCCATACGATTGATTCAGCACAACCCAGCTCTTCGCAACCTTATTGCCACAG 241
Db |||||||
QY 588 CTCAGGCCCAGAAATCAGCAGCAGACAGAGGTGTAAAACCTGAAGAGAGTGAACCTTTC 647
Db |||||||
QY 242 CTCAGGCCCAGAAATCAGCAGCAGACAGAGGTGTAAAACCTGAAGAGAGTGAACCTTTC 301
Db |||||||
QY 648 CCTCTGCGCCCTGGGTCACCTCTCTCCCTGATGACCTCTCTGCTTTAGATTGAAGATC 707
Db |||||||
QY 302 CCTCTGCGCCCTGGGTCACCTCTCTCCCTGATGACCTCTCTGCTTTAGATTGAAGATC 361
Db |||||||
QY 708 CCAATGCACCATTCAGATCCGGCAGTGCACCCAGAGAGTGCATTTTATCTGCGGAAG 767
Db |||||||
QY 362 CCAATGCACCATTCAGATCCGGCAGTGCACCCAGAGAGTGCATTTTATCTGCGGAAG 421
Db |||||||
QY 768 GGGAACTGTGACTGAACCTCAGCTGGCACTCTCTGCGCAGCTCTCTACCAAGGAGTGG 827
Db |||||||
QY 422 GGGAACTGTGACTGAACCTCAGCTGGCACTCTCTGCGCAGCTCTCTACCAAGGAGTGG 481
Db |||||||
```



Qy 828 CCACAAATCTGGCCCGCGGGCTTTGACTGTGCTAATGAGAGTGTCTGGAGACCCCTAA 887  
Db 482 CCACAAATCTGGCCCGCGGGCTTTGACTGTGCTAATGAGAGTGTCTGGAGACCCCTAA 541  
Qy 888 CTGATGGGCACATGAGTATGCTTAAAGTTTACCAGTTGCTGCGTTTTCGTTGGTGGACC 947  
Db 542 CTGATGGGCACATGAGTATGCTTAAAGTTTACCAGTTGCTGCGTTTTCGTTGGTGGACC 601  
Qy 948 GGGAGGCGCGGCTGGGACAGACTCTTTTCTGATGTGATGGAGAGGATTTCCATGAAG 1007  
Db 602 GGGAGGCGCGGCTGGGACAGACTCTTTTCTGATGTGATGGAGAGGATTTCCATGAAG 661  
Qy 1008 TGGGTATGGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCACCCTCAAGAGCTATC 1067  
Db 662 TGGGTATGGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCACCCTCAAGAGCTATC 721  
Qy 1068 ACAGTTACATGCTACAGATAGTAAGCACTCTCTGAAGATCAAGGAGGAACTGTGACATC 1127  
Db 722 ACAGTTACATGCTACAGATAGTAAGCACTCTCTGAAGATCAAGGAGGAACTGTGACATC 781  
Qy 1128 CTGAGAAGGCCACAGAGGAGCGCTTAAACCTGTGAAGATCAAGGAGGAACTGTGAGGAGCA 1187  
Db 782 CTGAGAAGGCCACAGAGGAGCGCTTAAACCTGTGAAGATCAAGGAGGAACTGTGAGGAGCA 841  
Qy 1188 TCACTTTTCTGTCAGTGAGGAGCTGGAGGCTGACCTTGCTTCTGGAGCAGCTCACTGC 1247  
Db 842 TCACTTTTCTGTCAGTGAGGAGCTGGAGGCTGACCTTGCTTCTGGAGCAGCTCACTGC 901  
Qy 1248 CTATGGAGTGCTTGGGGCTCAGAGCAAGGCTTCCCATCAACCTGGAGGTTGAAGCTT 1307  
Db 902 CTATGGAGTGCTTGGGGCTCAGAGCAAGGCTTCCCATCAACCTGGAGGTTGAAGCTT 961  
Qy 1308 CACCACAGGCTTCAAGTGCAGAGGTAAGTCTCTCTCTTCTGGAATCTGGCCCATGTGA 1367  
Db 962 CACCACAGGCTTCAAGTGCAGAGGTAAGTCTCTCTCTTCTGGAATCTGGCCCATGTGA 1021  
Qy 1368 AATGAGGCTTCAAGAAAGTGAAGAGGCAATCTCTGGGCATGTGCTGGGCAGTG 1427  
Db 1022 AATGAGGCTTCAAGAAAGTGAAGAGGCAATCTCTGGGCATGTGCTGGGCAGTG 1081  
Qy 1428 ATGCTCTCGAGGAGCTATGTCAGGATGAGTGAAGCTGCTGGGCATGTGCTGGGCAGTG 1487  
Db 1082 ATGCTCTCGAGGAGCTATGTCAGGATGAGTGAAGCTGCTGGGCATGTGCTGGGCAGTG 1141  
Qy 1488 ACTCAGATAGCAGCTATGTTCCACTTCCACTCAGACGCTCATGGGCTCTCAGAGCCCTGATG 1547  
Db 1142 ACTCAGATAGCAGCTATGTTCCACTTCCACTCAGACGCTCATGGGCTCTCAGAGCCCTGATG 1201  
Qy 1548 TCAACAGGCTGCAAGAGAGGATGAGGAAATATAA 1585  
Db 1202 TCAACAGGCTGCAAGAGAGGATGAGGAAATATAA 1239

## RESULT 3

US-09-922-261-80  
; Sequence 80, Application US/0992261  
; Patent No. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US/09/461,697  
; PRIOR FILING DATE: 1999-12-14

; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 80  
; LENGTH: 1092  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-922-261-80

Query Match 63.8%; Score 1092; DB 11; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 7.1e-312; Indels 0; Gaps 0;  
Matches 1092; Conservative 0; Mismatches 0

Qy 494 ATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGATTCAGTTGATTGAG 553  
Db 1 ATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGATTCAGTTGATTGAG 60  
Qy 554 CACAACCGAGCTCTTCGCAACCTTATGCGCAGAGCTCAGCCCGAGATCAGCAGCAGACA 613  
Db 61 CACAACCGAGCTCTTCGCAACCTTATGCGCAGAGCTCAGCCCGAGATCAGCAGCAGACA 120  
Qy 614 GAAGGTGTAAAAAAGTGAAGAGAGTGAACCTTCTCCCTCGTCCCTGGGTCCCTCCTCTC 673  
Db 121 GAAGGTGTAAAAAAGTGAAGAGAGTGAACCTTCTCCCTCGTCCCTGGGTCCCTCCTC 180  
Qy 674 CTTGATGACCTCTCTGCTTTAGATGTGAAGATCCCAATGCCAATTCAGATTCAGGAC 733  
Db 181 CTTGATGACCTCTCTGCTTTAGATGTGAAGATCCCAATGCCAATTCAGATTCAGGAC 240  
Qy 734 AGTGACCCAGAGAGTACCTTTTATCGTGGGAAAGGGGAACTGTGACTGACTGACTGG 793  
Db 241 AGTGACCCAGAGAGTACCTTTTATCGTGGGAAAGGGGAACTGTGACTGACTGACTGG 300  
Qy 794 CACTCTGTCTGGCAGCTCTCTTACCAGGAGTGGCCACAACTCTGGCCACCGGGCTTT 853  
Db 301 CACTCTGTCTGGCAGCTCTCTTACCAGGAGTGGCCACAACTCTGGCCACCGGGCTTT 360  
Qy 854 GACTGTCTAATGAGAGTGTCTTGGAGACCCCTAACTGATGTGGCAGATGATGATTCCT 913  
Db 361 GACTGTCTAATGAGAGTGTCTTGGAGACCCCTAACTGATGTGGCAGATGATGATTCCT 420  
Qy 914 AAGTTTACCAAGTGTCTGCGTTTGTCTGGACCGGGAGCCCGGGTGGCAGAGCTCT 973  
Db 421 AAGTTTACCAAGTGTCTGCGTTTGTCTGGACCGGGAGCCCGGGTGGCAGAGCTCT 480  
Qy 974 TTTCTCTGATGTGATGGAGAGTATTCATGAAGTGGGTATTCGAGTGTGCTCTCCCTC 1033  
Db 481 TTTCTCTGATGTGATGGAGAGTATTCATGAAGTGGGTATTCGAGTGTGCTCTCCCTC 540  
Qy 1034 CAGAAGTCTTGGCAGCACCCTATCAAGGAGTATCAAGTGTATGATGATGATGATGAG 1093  
Db 541 CAGAAGTCTTGGCAGCACCCTATCAAGGAGTATCAAGTGTATGATGATGATGATGAG 600  
Qy 1094 CAACTCTCTGAAGATATGAAGGATTTCAATCTCTGAGAGCCACAGAGGAGCTTAA 1153  
Db 601 CAACTCTCTGAAGATATGAAGGATTTCAATCTCTGAGAGCCACAGAGGAGCTTAA 660  
Qy 1154 CTTGTGAAGATCAAGGAGGAACTGTGAGGAGCATCACCTTTCTCTGCTGAGGAGCTG 1213  
Db 661 CTTGTGAAGATCAAGGAGGAACTGTGAGGAGCATCACCTTTCTCTGCTGAGGAGCTG 720  
Qy 1214 GAGGCTGACCTTCTTCTGGAGACCTACTGCTATGGAGTGTGCTGGGCTCAGAGC 1273  
Db 721 GAGGCTGACCTTCTTCTGGAGACCTACTGCTATGGAGTGTGCTGGGCTCAGAGC 780  
Qy 1274 GAAGCTTCTCCATCTAACCTGGAGGTTGAAGTTCACCAGAGCTTCAAGTGCAGAGGTA 1333  
Db 781 GAAGCTTCTCCATCTAACCTGGAGGTTGAAGTTCACCAGAGCTTCAAGTGCAGAGGTA 840  
Qy 1334 AATGCTTCTCTCTTGGATCTGGCCCATGTGAAGTGGAGCTCAAGAAAGTGAAGAA 1393  
Db 841 AATGCTTCTCTCTTGGATCTGGCCCATGTGAAGTGGAGCTCAAGAAAGTGAAGAA 900  
Qy 1394 GGCAATGTCTCTGGGATGCTGTGGGAGTGTGCTGGGAGTGTGCTTTCAGGAGGCTATGTCAGGC 1453

```
Db 901 GCAATGTCCTGGGCATGGTGTGCTGGCAGTGTATGCTTCGAGGAGCCTATGTCAGGC 960
QY 1454 ATGAGTGAAGTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGTTGCCAC 1513
Db 961 ATGAGTGAAGTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGTTGCCAC 1020
QY 1514 TCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATG 1573
Db 1021 TCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATG 1080
QY 1574 AGGAAATATATA 1585
Db 1081 AGGAAATATATA 1092

RESULT 4
US-09-922-261-92
; Sequence 92, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-92
```

```
Query Match 35.1%; Score 600; DB 11; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.2e-166;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 ATGGAGCAGGTATTCATGAAGTGGGTATGCGAGTGTGCTCCCTCCAGAAAGTTCTGG 1045
Db 1 ATGGAGCAGGTATTCATGAAGTGGGTATGCGAGTGTGCTCCCTCCAGAAAGTTCTGG 60

QY 1046 CAGCACCCATCAGGACTATCAGATTACATGCTACAGATTAGTAAGCACTCTCTGAA 1105
Db 61 CAGCACCCATCAGGACTATCAGATTACATGCTACAGATTAGTAAGCACTCTCTGAA 120

QY 1106 GAATATGAAGGATTTCAATCCTGAGAGGCCACAGAGGCGGTAAACCTGTGAAGATC 1165
Db 121 GAATATGAAGGATTTCAATCCTGAGAGGCCACAGAGGCGGTAAACCTGTGAAGATC 180

QY 1166 AAGGAGAACCTGTGAGCGCATCATCTTCTGTGCTAGTGGAGCTGGAGCGTGACCTT 1225
Db 181 AAGGAGAACCTGTGAGCGCATCATCTTCTGTGCTAGTGGAGCTGGAGCGTGACCTT 240

QY 1226 GCTTCTGGAGACCACTGCTGCTATGGAGTGTGTTGGGCTCAGAGCGAAGCGTTCCCA 1285
Db 241 GCTTCTGGAGACCACTGCTGCTATGGAGTGTGTTGGGCTCAGAGCGAAGCGTTCCCA 300

QY 1286 TCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAGTCAGAGGTTAAATGCTTCTCT 1345
Db 301 TCTAACCTGGAGGTTGAAGCTTCAACACAGGCTTCAAGTCAGAGGTTAAATGCTTCTCT 360

QY 1346 CTTTGGATCTGCGCCATGTGAATGGAGCCTCAAGAAAGTGAAGAGGCAATGTCTCT 1405
```

```
Db 361 CTTTGGAAATCTGGCCCATGTGAAATGGAGCTCAAGAAAGTGAAGAAAGCAATGTCTCT 420
QY 1406 GGGCATGTGTGCTGGGCGAGTGTCTTCGAGGAGCCCTATGTCAGGCGATGAGTGAAGCT 1465
Db 421 GGGCATGTGTGCTGGGCGAGTGTCTTCGAGGAGCCCTATGTCAGGCGATGAGTGAAGCT 480
QY 1466 GGGATTCTTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGACAGC 1525
Db 481 GGGATTCTTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCCACTCCACTGACAGC 540
QY 1526 CTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATGAGGAAATATATA 1585
Db 541 CTCATGGGTCCTCCCTGTTTCAACACGCGTGCAGAAAGAGGATGAGGAAATATATA 600

RESULT 5
US-09-922-261-96
; Sequence 96, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-96
```

```
Query Match 29.8%; Score 510; DB 11; Length 510;
Best Local Similarity 100.0%; Pred. No. 4.5e-140;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1076 ATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTTGTCAATCCTGAGA 1135
Db 1 ATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTTGTCAATCCTGAGA 60

QY 1136 GCCACAGAGGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGCGACATCACTTTT 1195
Db 61 GCCACAGAGGCGCTAAACCTGTGAAGATCAAGGAGGAACCTGTGAGCGACATCACTTTT 120

QY 1196 CCTGTCACTGAGGAGCTGGAGCGTCACTTCTTCTGAGAGCACCTGCTCCTATGGGA 1255
Db 121 CCTGTCACTGAGGAGCTGGAGCGTCACTTCTTCTGAGAGCACCTGCTCCTATGGGA 180

QY 1256 GTGCTTGGGCTCAGAGGAGCGCTTCCCATCTACCTGGAGGTTGAAGCTTCACACAG 1315
Db 181 GTGCTTGGGCTCAGAGGAGCGCTTCCCATCTACCTGGAGGTTGAAGCTTCACACAG 240

QY 1316 GCTTCAAGTGCAGAGGTTAAATGCTTCTCTTTTGAATCTGGCCCATGTGAAATGGAG 1375
Db 241 GCTTCAAGTGCAGAGGTTAAATGCTTCTCTTTTGAATCTGGCCCATGTGAAATGGAG 300

QY 1376 CCTCAAGAAAGTGAAGAGGCAATGTCTCTGGGCATGTTGCTGGGCGAGTGTCTTC 1435
Db 301 CCTCAAGAAAGTGAAGAGGCAATGTCTCTGGGCATGTTGCTGGGCGAGTGTCTTC 360

QY 1436 GAGGAGCCTATGTGAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGTGACTCAGAT 1495
```



; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31390  
; LENGTH: 491  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(491)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31390

Query Match  
Best Local Similarity 19.7%; Score 337; DB 12; Length 491;  
Matches 346; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
  
QY 1 ACGGATCTTCCTCAGGCTCTCGAGGTCCAGACGCCGCCAGCCGCTCTGCGAGC 60  
Db 70 ACGGAGCCTTCCTCAGGCTCTCGAGGTCCAGACTGCCGCCAGTCCGCTCTGCGAGC 129  
  
QY 61 CAGCAGTGAATAGTGTGTACTCTCTCTCGGTTCAGGTCCAGACCTCCCGCTCTTCC 120  
Db 130 CACCACTGAATAGTGTGTACTCTCTCTCGGTTCAGGTCCAGACCTCCCGCTCTTCC 189  
  
QY 121 GCTCCTCTGAACGTCAGGACCTCAGACCTGTGATGGCGCTCGCGCGGCGACC 180  
Db 190 GCTCCTCTGAACGTCAGGACCTCAGACCTGTGATGGCGCTCGCGCGGCGACC 249  
  
QY 181 GTGACCGAGGAACCCCTGGAGGACTTGGGCTCTTGGCTCCGCTGTCTTCG 240  
Db 250 GTGACCGAGGAACCCCTGGAGGACTTGGGCTCTTGGCTCCGCTGTCTTCG 309  
  
QY 241 TGCTCCTTCGCGGCAAGATCTCATTATCAGTCTTTGACCGACACAGAAATGCTGGCA 300  
Db 310 CGCTCCTTCGCGGCAAGATCTCATTATCAGTCTTTGACCGACACAGAAATGCTGGCA 369  
  
QY 301 TTGTATAAATGTTGTTGAATCTGAAGACATATGGAACAATCTGCAAGATACTG 360  
Db 370 TTGTATAAATGTTGTTGAATCTGAAGACATATGGAACAATCTGCAAGATACTG 429  
  
QY 361 G 361  
Db 430 G 430

RESULT 9  
US-09-922-261-100  
; Sequence 100, Application US/09922261  
; Patent No. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Purnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US/09/461,697  
; PRIOR FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 100  
; LENGTH: 336  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-922-261-100

Query Match 19.6%; Score 336; DB 11; Length 336;

Best Local Similarity 100.0%; Pred. No. 9.3e-89;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1250 ATGGAGTGTCTGGGCTCAGAGCGAAGCGTCCCATCTACCTTGGAGTTGAAGCTTCA 1309  
Db 1 ATGGAGTGTCTGGGCTCAGAGCGAAGCGTCCCATCTACCTTGGAGTTGAAGCTTCA 60  
  
QY 1310 CCACAGGCTTCAAGTGCAGAGGTAAATGCTTCTCTCTTTTGAATCTGGCCCATGTGAAA 1369  
Db 61 CCACAGGCTTCAAGTGCAGAGGTAAATGCTTCTCTCTTTTGAATCTGGCCCATGTGAAA 120  
  
QY 1370 ATGGAGCCTCAAGAAAGTGAAGAGGCAATCTCTCTGGGCATGTGTGTCGGCAGTGAT 1429  
Db 121 ATGGAGCCTCAAGAAAGTGAAGAGGCAATCTCTCTGGGCATGTGTGTCGGCAGTGAT 180  
  
QY 1430 GTCCTCAGGAGCCTATGTCTAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGAC 1489  
Db 181 GTCCTCAGGAGCCTATGTCTAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGAC 240  
  
QY 1490 TCAGATAGCAGCTATGTTCCCTCCACTCCACTCAGAGCCTCATGGGTCTCTCCCTGTTTC 1549  
Db 241 TCAGATAGCAGCTATGTTCCCTCCACTCCACTCAGAGCCTCATGGGTCTCTCCCTGTTTC 300  
  
QY 1550 AACGAGCGCTCAAGAAAGAGAGATGAGCAAAATATAA 1585  
Db 301 AACGAGCGCTCAAGAAAGAGAGATGAGCAAAATATAA 336

RESULT 10  
US-09-764-847-352  
; Sequence 352, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 352  
; LENGTH: 517  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (59)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (482)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (508)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-847-352

Query Match 17.2%; Score 293.8; DB 11; Length 517;  
Best Local Similarity 96.1%; Pred. No. 3.5e-76;  
Matches 321; Conservative 1; Mismatches 9; Indels 3; Gaps 2;  
  
QY 21 CTCTCGAGGTCCAGACAGCGCCCGCCGCTCTCGCAGCAGCAGTGAATAGTGTGTA 80  
Db 17 CCCACGGCTCCGACAGCGCCCGCCGCTCTCGCAGCAGCAGTGAATAGTGTGTA 76  
  
QY 81 CCTCCTTGTCTCGGTTTCAGGTTCAGAGCTCCCTCTCTTCGGGTGCGCTGAACCTCAGGC 140  
Db 77 CCTCCTTGTCTCGGTTTCAGGTTCAGAGCTCCCTCTCTTCGGGTGCGCTGAACCTCAGGC 136  
  
QY 141 GACCTCAGGACCTGTGATTGGCCCTTGGCCCGGCGGACCTGACCGAGGAAACCCCTGG 200  
Db 137 GACCTCAGGACCTGTGATTGGCCCTTGGCG--GRCGACCGGTGACCGAGGAAACCCCTGG 194

RESULT 12  
US-09-918-995-3654

RESULT 13  
US-09-922-261-104  
; Sequence 104, Application US/09922261  
; Patent NO. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US/09/461,697  
; PRIOR FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 104  
; LENGTH: 216  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-922-261-104

Query Match 12.6%; Score 216; DB 11; Length 216;  
Best Local Similarity 100.0%; Pred. No. 2.1e-53;  
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 1 ATGGAGCCTCAGAAAGTCAAGAGCAATGCTCTGGCATGTGTCTGGCAGTGAT 60
QY 1430 GTCTTCGAGGAGCCTATCTCAGGATGAGTGAAGCTGGATTCTTCAGAGCCCTGATGAC 1489
Db 61 GTCTTCGAGGAGCCTATCTCAGGATGAGTGAAGCTGGATTCTTCAGAGCCCTGATGAC 120
QY 1490 TCAGATAGCAGTATGTTCCACTCCACTGACAGCCCTCATGGGTCCTCCCTGTTTC 1549
Db 121 TCAGATAGCAGTATGTTCCACTCCACTGACAGCCCTCATGGGTCCTCCCTGTTTC 180
QY 1550 AACAGCCGCTCAAGAGAGGATGAGGAAATATAA 1585
Db 181 AACAGCCGCTCAAGAGAGGATGAGGAAATATAA 216
```

```
RESULT 14
US-09-922-261-94
; Sequence 94, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-94
```

```
Query Match 9.3%; Score 159; DB 11; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.2e-36;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1002 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAAAGTCTGGCAGCAGCCGATCAAGG 1061
Db 1 ATGAAGTGGGTATTGGCAGTGTCTCTCCCTCCAGAAAGTCTGGCAGCAGCCGATCAAGG 60

QY 1062 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTG 1121
Db 61 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTG 120

QY 1122 TCAATCTCTGAGAAGCCACAGGAGCGCTAAACCTGTGA 1160
Db 121 TCAATCTCTGAGAAGCCACAGGAGCGCTAAACCTGTGA 159
```

```
RESULT 15
US-09-922-261-112
; Sequence 112, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112
; LENGTH: 141
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-261-112
```

```
Query Match 8.2%; Score 141; DB 11; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.4e-31;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1445 ATGTCAGGCATGAGTGAAGCTGGGATTCTTCAGAGCCCTGATGACTCAGATAGCAGCTAT 1504
Db 1 ATGTCAGGCATGAGTGAAGCTGGGATTCTTCAGAGCCCTGATGACTCAGATAGCAGCTAT 60

QY 1505 GGTTCCTCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTCACACCGCGCTGCAAG 1564
Db 61 GGTTCCTCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTCACACCGCGCTGCAAG 120

QY 1565 AAGAGGATGAGGAAATATAA 1585
Db 121 AAGAGGATGAGGAAATATAA 141
```

Search completed: July 18, 2003, 09:55:19  
Job time : 403 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:40:42 ; Search time 89 Seconds  
(without alignments)  
5895.775 Million cell updates/sec

Title: US-09-857-308-2

Perfect score: 1711

Sequence: 1 acgcgatccttccttcaggc.....aaaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2.6/prodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2.6/prodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2.6/prodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2.6/prodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2.6/prodata/1/ina/PCUS\_COMB.seq.\*
- 6: /cgn2.6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	75.3	1825	4	US-09-461-697-75
2	1233.2	72.1	1239	4	US-09-461-697-76
3	1092	63.8	1092	4	US-09-461-697-80
4	600	35.1	600	4	US-09-461-697-92
5	510	29.8	510	4	US-09-461-697-96
6	336	19.6	336	4	US-09-461-697-100
7	216	12.6	216	4	US-09-461-697-104
8	159	9.3	159	4	US-09-461-697-94
9	141	8.2	141	4	US-09-461-697-112
10	132	7.7	132	4	US-09-461-697-114
11	105	6.1	105	4	US-09-461-697-78
12	96	5.6	96	4	US-09-461-697-88
13	90	5.3	90	4	US-09-461-697-116
14	87	5.1	87	4	US-09-461-697-90
15	72	4.2	72	4	US-09-461-697-118
16	69	4.0	69	4	US-09-461-697-84
17	60	3.5	60	4	US-09-461-697-106
18	57	3.3	57	4	US-09-461-697-120
19	51	3.0	51	4	US-09-461-697-98
20	48.4	2.8	2502	3	US-09-234-332-1
21	48	2.8	48	4	US-09-461-697-108
22	43.6	2.5	2089	1	US-08-552-142A-1
23	43.6	2.5	2089	1	US-08-910-973-1
24	43.6	2.5	2089	4	US-09-499-227-1
25	43.6	2.5	2089	5	PCR-US95-05741-1
26	39	2.3	45	4	US-09-461-697-124
27	38.4	2.2	1831	4	US-09-336-536-15

28	38	2.2	1821	4	US-09-149-476-90	Sequence 90, Appl	
c	37.4	2.2	7218	1	US-08-232-463-14	Sequence 14, Appl	
	36.6	2.1	18443	4	US-09-078-294-6	Sequence 6, Appl	
	36.4	2.1	2525	4	US-09-234-393-39	Sequence 39, Appl	
	32	36.4	2.1	2525	4	US-09-865-171-39	Sequence 39, Appl
33	36.2	2.1	2601	4	US-08-569-749-3	Sequence 3, Appl	
34	36.2	2.1	2601	5	PCT-US96-12860-3	Sequence 3, Appl	
35	36	2.1	1296	5	PCT-US94-12912-3	Sequence 3, Appl	
36	36	2.1	1571	5	PCT-US94-12912-2	Sequence 2, Appl	
37	36	2.1	2025	4	US-09-149-476-316	Sequence 316, App	
38	36	2.1	2346	4	US-09-149-476-193	Sequence 193, App	
39	36	2.1	3204	4	US-09-336-643A-26	Sequence 26, Appl	
40	35.6	2.1	1400	1	US-08-085-122-12	Sequence 12, Appl	
c	41	35.6	2.1	1400	4	US-08-442-108B-22	Sequence 22, Appl
	42	35.6	2.1	1770	4	US-08-943-731-146	Sequence 146, App
	43	35.6	2.1	2301	4	US-09-232-191-8	Sequence 8, Appl
	44	35.6	2.1	2301	4	US-09-232-200-8	Sequence 8, Appl
45	35.6	2.1	2301	4	US-09-232-197-8	Sequence 8, Appl	

ALIGNMENTS

RESULT 1  
US-09-461-697-75  
; Sequence 75, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Purnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 75  
; LENGTH: 1825  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-75

Query Match	75.3%	Score 1288;	DB 4;	Length 1825;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1291;	Conservative	0;	Mismatches	5;
Indels	0;	Gaps	0;	
QY	344	AAATCTGCAAGATACTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC	403	
Db	26	AAATCTGCAAGATACTGGGAGAGATACCAATATCATCAAGCCAGACCAAGAGTTCC	85	
QY	404	TTTCGATTTCTCCACGGGAGTTCCGTCGTGGTGGAGTCCATGACCCACCCCTGCACCAA	463	
Db	86	TTTCGATTTCTCCACGGGAGTTCCGTCGTGGTGGAGTCCATGACCCACCCCTGCACCAA	145	
QY	464	CCCTCAGCCCAACGCGGAGCCCTCCACTATGCTGGACATCCCTTCAGAGCATGTAGT	523	
Db	146	CCCTCAGCCCAACGCGGAGCCCTCCACTATGCTGGACATCCCTTCAGAGCATGTAGT	205	
QY	524	CTCACCATCCATACGATTCAGTTGATTCAGCACACCGAGCTCTTCGGCAACCTTATGTC	583	
Db	206	CTCACCATCCATACGATTCAGTTGATTCAGCACACCGAGCTCTTCGGCAACCTTATGTC	265	
QY	584	ACAGCTCAGGCCCCAGAAATCAGCAGACAGAGAGTGTAAAAAAGTGAAGAGAGTGAACCT	643	
Db	266	ACAGCTCAGGCCCCAGAAATCAGCAGACAGAGAGTGTAAAAAAGTGAAGAGAGTGAACCT	325	



QY 644 CTTCCCTCGTCCCTGGGTACCTCCCTCCTGATGACCTCCCTGCTTTAGATTGTAAG 703  
Db CTTCCCTCGTCCCTGGGTACCTCCCTCCTGATGACCTCCCTGCTTTAGATTGTAAG 385  
QY 704 AATCCCAATGACCATTCACATCCGACAGTCCGACAGTCCGACAGTCTTTATCGTGGG 763  
Db AATCCCAATGACCATTCACATCCGACAGTCCGACAGTCCGACAGTCTTTATCGTGGG 445  
QY 764 AAAGGGAACTGTGACTGAATCACTGAGTGGCAGTCCCTGCGGACGCTCTCTACAGGCA 823  
Db AAAGGGAACTGTGACTGAATCACTGAGTGGCAGTCCCTGCGGACGCTCTCTACAGGCA 505  
QY 824 GTGGCCACATCTTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTCTCTGGAGACC 883  
Db GTGGCCACATCTTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTCTCTGGAGACC 565  
QY 884 CTAAGTGTGGCAGATGAGTATGCTTAAGTTTACCAAGTGTGCTTTGCTGTG 943  
Db CTAAGTGTGGCAGATGAGTATGCTTAAGTTTACCAAGTGTGCTTTGCTGTG 625  
QY 944 GACCGGAGGCGCGCTGGGACAGTCTTTTCTGATGTGATGGAGGATTTCCAT 1003  
Db GACCGGAGGCGCGCTGGGACAGTCTTTTCTGATGTGATGGAGGATTTCCAT 685  
QY 1004 GAAGTGGGTATGTCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGAC 1063  
Db GAAGTGGGTATGTCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGAC 745  
QY 1064 TATCAGAGTTACATGCTACAGATTAGTAAGCAACTCTGAGAAATATGAAGATTGTC 1123  
Db TATCAGAGTTACATGCTACAGATTAGTAAGCAACTCTGAGAAATATGAAGATTGTC 805  
QY 1124 AATCCTGAGAGGCGCACAGAGCGCTAAAGCTCTGAGAGTCAAGGAGGAACTGTGAGC 1183  
Db AATCCTGAGAGGCGCACAGAGCGCTAAAGCTCTGAGAGTCAAGGAGGAACTGTGAGC 865  
QY 1184 GACATCACTTTTCTGTCAGTGGAGCTGTGAGCTGACCTTGTCTGAGAGCAGTCA 1243  
Db GACATCACTTTTCTGTCAGTGGAGGCTGTGAGCTGACCTTGTCTGAGAGCAGTCA 925  
QY 1244 CTGCTATGGAGTGTCTGGGCTCAGAGCAAGCTTCCATCTAAGCTGGAGGTTGAA 1303  
Db CTGCTATGGAGTGTCTGGGCTCAGAGCAAGCTTCCATCTAAGCTGGAGGTTGAA 985  
QY 1304 GCTTCACACAGGCTTCAAGTGCAGAGTAAATGCTTCTCTCTTTTGAATCTGGCCCAT 1363  
Db GCTTCACACAGGCTTCAAGTGCAGAGTAAATGCTTCTCTCTTTTGAATCTGGCCCAT 1045  
QY 1364 GTGAAATGGAGCTCAAGAAAGTGAAGAGGCAATGTCTGCGGCAATGTGCTGGGC 1423  
Db GTGAAATGGAGCTCAAGAAAGTGAAGAGGCAATGTCTGCGGCAATGTGCTGGGC 1105  
QY 1424 AGTATGTCTTCGAGGAGCTTATCTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCT 1483  
Db AGTATGTCTTCGAGGAGCTTATCTCAGGATGAGTGAAGTGGATTCCTCAGAGCCCT 1165  
QY 1484 GATGACTCAGATAGCAGTATGTTTCCACTTCCACTGACAGCTCTATGGGCTCTCCCT 1543  
Db GATGACTCAGATAGCAGTATGTTTCCACTTCCACTGACAGCTCTATGGGCTCTCCCT 1225  
QY 1544 GTTTTCAACAGGCTCGAAGAGGATGAGGAAATATATAAGGAAAGAGAGGAGATG 1603  
Db GTTTTCAACAGGCTCGAAGAGGATGAGGAAATATATAAGGAAAGAGAGGAGATG 1285  
QY 1604 TTTTGTCCAGACCTACTAGACCCCAACAGAAAGTT 1639  
Db TTTTGTCCAGACCTACTAGACCCCAACAGAAAGTT 1321

RESULT 2  
US-09-461-697-76  
; Sequence 76, Application US/09461697  
; Patent No. 6277974

GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 76  
; LENGTH: 1239  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-76

Query Match 72.18; Score 1233.2; DB 4; Length 1239;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 348 TGAAGATATCTGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCTTCG 407  
Db 2 TGTGAGATCTGGGAGAGATACCAATATCATCAAGCCAGACCAACAGAGTTCTTCG 61  
QY 408 ATTGTCTCCACGGGAGTTCCGTCTGCTGGAAGTCCATGACCCACCCCTCGACCAACCT 467  
Db 62 ATTGTCTCCACGGGAGTTCCGTCTGCTGGAAGTCCATGACCCACCCCTCGACCAACCT 121  
QY 468 CAGCAACAAGCCGAGCCGCCACTATGCTGGACATCCCTCAGAGCCATGATGTCTCA 527  
Db 122 CAGCAACAAGCCGAGCCGCCACTATGCTGGACATCCCTCAGAGCCATGATGTCTCA 181  
QY 528 CCATCCATAGATTCAGTTGATTGATTCAGCACCAACCGAGCTCTTCCGCAACCTATTGCCACAG 587  
Db 182 CCATCCATAGATTCAGTTGATTGATTCAGCACCAACCGAGCTCTTCCGCAACCTATTGCCACAG 241  
QY 588 CTAGGCCCCAGATTCAGCACAGACAGAGTGTAAACTGAAGAGTGAACCTCTTC 647  
Db 242 CTAGGCCCCAGATTCAGCACAGACAGAGTGTAAACTGAAGAGTGAACCTCTTC 301  
QY 648 CCTCGTCCCTGGTCCACCTCTCTCCCTGATGACCTCTCTCCCTTTAGATTGTAAGATC 707  
Db 302 CCTCGTCCCTGGTCCACCTCTCTCCCTGATGACCTCTCTCCCTTTAGATTGTAAGATC 361  
QY 708 CCAATGCACCATTCAGATCCGACAGTCCGACAGTCCGACAGAGTGAACCTTTATCGTGGGAAAG 767  
Db 362 CCAATGCACCATTCAGATCCGACAGTCCGACAGTCCGACAGAGTGAACCTTTATCGTGGGAAAG 421  
QY 768 GGGAACTGTGACTGAATCAGCTGGACCTCTCTGTCGGAGCTCTCTACAGGAGTGG 827  
Db 422 GGGAACTGTGACTGAATCAGCTGGACCTCTCTGTCGGAGCTCTCTACAGGAGTGG 481  
QY 828 CCACATCTCTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCGAGACCTTAA 887  
Db 482 CCACATCTCTGGCCAGCGGGCTTTGACTGTGCTAATGAGAGTGTCTCGAGACCTTAA 541  
QY 888 CTGATGTGGCAGATGAGTATTGCTTAAAGTTTACCAGTTTCTGCTGCTTTGCTGGACC 947  
Db 542 CTGATGTGGCAGATGAGTATTGCTTAAAGTTTACCAGTTTCTGCTGCTTTGCTGGACC 601  
QY 948 GGGAGGCCGCTGGGACAGACTCTTTCTGATGTGATGGAGGAGTATTCATGAAG 1007  
Db 602 GGGAGGCCGCTGGGACAGACTCTTTCTGATGTGATGGAGGAGTATTCATGAAG 661  
QY 1008 TGGGTATTGGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGACTATC 1067  
Db 662 TGGGTATTGGCAGTGTCTCTCCCTCCAGAAAGTTCTGGCAGCAGCCGATCAAGGACTATC 721



QY 1068 ACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTCAATC 1127  
D 1127 ACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAATATGAAGGATTGTCAATC 781  
QY 1128 CTGAGAGGCCACAGAGGAGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGGAGCA 1187  
D 782 CTGAGAGGCCACAGAGGAGCTAAACCTGTGAAGATCAAGGAGAACCTGTGAGGAGCA 841  
QY 1188 TCACTTTTCTGTCAGTGAGGAGCTGAGGCTGACCTTCTCTGGAGACCACTCACTGC 1247  
D 842 TCACTTTTCTGTCAGTGAGGAGCTGAGGCTGACCTTCTCTGGAGACCACTCACTGC 901  
QY 1248 CTATGGAGTGTCTGGGGCTCAGAGCGAAGCGCTTCCCATCTAACCTGGAGGTTGAAGCTT 1307  
D 902 CTATGGAGTGTCTGGGGCTCAGAGCGAAGCGCTTCCCATCTAACCTGGAGGTTGAAGCTT 961  
QY 1308 CACCACAGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTGGAAATCTGGCCCATGTGA 1367  
D 962 CACCACAGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTGGAAATCTGGCCCATGTGA 1021  
QY 1368 AAATGAGGCTCAAGAAAGTGAAGAGCAATGCTCTGGGATGCTGTGGGCAAGT 1427  
D 1022 AAATGAGGCTCAAGAAAGTGAAGAGCAATGCTCTGGGATGCTGTGGGCAAGT 1081  
QY 1428 ATGCTTTCAGAGGAGCTATGTCAAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATG 1487  
D 1082 ATGCTTTCAGAGGAGCTATGTCAAGGATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATG 1141  
QY 1488 ACTCAGATAGCAGTATGTTCCCACTCCACTGCAGAGCCCTCATGGGGTCTCCCTGTTT 1547  
D 1142 ACTCAGATAGCAGTATGTTCCCACTCCACTGCAGAGCCCTCATGGGGTCTCCCTGTTT 1201  
QY 1548 TCAACACGCGCTGCAAGAGAGGATGAGGAAATATAA 1585  
D 1202 TCAACACGCGCTGCAAGAGAGGATGAGGAAATATAA 1239

RESULT 3

US-09-461-697-80  
; Sequence 80, Application US/09461697  
; Patent No. 627974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Purnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 80  
; LENGTH: 1092  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-80

Query Match 63.8%; Score 1092; DB 4; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 7e-302;  
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 494 ATGCTGGACATCCCTCAGAGGAGCTAGTCTCACCATCCATACGATTCAGTTGATTCAG 553  
D 1 ATGCTGGACATCCCTCAGAGGAGCTAGTCTCACCATCCATACGATTCAGTTGATTCAG 60  
QY 554 CACAACCGAGCTCTTCGCAACCTTATGTCACAGCTCAGGCCCAAGTAATCAGCAGCAGCA 613

61 CACAACCGAGCTCTTCGCAACCTTATGTCACAGCTCAGGCCCAAGTAATCAGCAGCAGCA 120  
614 GAAGGTGTAAAACACTCAAGAGAGTGAACCTCTTCCTCGTGGCTGGGTCACTCTCTCTC 673  
121 GAAGGTGTAAAACACTCAAGAGAGTGAACCTCTTCCTCGTGGCTGGGTCACTCTCTCTC 180  
674 CCGTATGACCTCTCTGCTTTAGATTGTAAGAATCCCAATSCACCAATCTCCGACATCCGAC 733  
181 CCGTATGACCTCTCTGCTTTAGATTGTAAGAATCCCAATSCACCAATCTCCGACATCCGAC 240  
734 ATGACCCACAGAGAGTACTTTTATCGTGGAAAGGGGAACCTGTGACTGAATCAGTCTGAG 793  
241 ATGACCCACAGAGAGTACTTTTATCGTGGAAAGGGGAACCTGTGACTGAATCAGTCTGAG 300  
794 CACTCTCTGCGCAGCTCTCTACAGAGCAGTGGCCACAATCTCCGACATCCGACATCCGAC 853  
301 CACTCTCTGCGCAGCTCTCTACAGAGCAGTGGCCACAATCTCCGACATCCGACATCCGAC 360  
854 GACTGTGCTTAATGAGAGTCTCTGGAGACCTTAACCTGATGTGGCACAATGATGATGATG 913  
361 GACTGTGCTTAATGAGAGTCTCTGGAGACCTTAACCTGATGTGGCACAATGATGATGATG 420  
914 AAGTTTACCAAGTGTGCTGCTTTGCTGTGGACCGGGAGGCCGCTGGGACAGACTCTCT 973  
421 AAGTTTACCAAGTGTGCTGCTTTGCTGTGGACCGGGAGGCCGCTGGGACAGACTCTCT 480  
974 TTTCTCTGATGTGATGAGCAGCTATTCCATGAAGTGGGTATTGGCAGTGTGCTCTCCCTC 1033  
481 TTTCTCTGATGTGATGAGCAGCTATTCCATGAAGTGGGTATTGGCAGTGTGCTCTCCCTC 540  
1034 CAGAAGTGTCTGGCAGCACCAGCATCAAGGACTATCACAGTTACATGCTCAGATTAAGTAAG 1093  
541 CAGAAGTGTCTGGCAGCACCAGCATCAAGGACTATCACAGTTACATGCTCAGATTAAGTAAG 600  
1094 CAACTCTCTGAAGAATATGAAGAAGTGTCAATCTCTGAGAAGGCCACAGAGAACGCTAAA 1153  
601 CAACTCTCTGAAGAATATGAAGAAGTGTCAATCTCTGAGAAGGCCACAGAGAACGCTAAA 660  
1154 CCTGTGAGATCAAGAGGAGAACCTGTGAGCAGCATCACTTTTCTCTGCTCAGTCAGGAGCTG 1213  
661 CCTGTGAGATCAAGAGGAGAACCTGTGAGCAGCATCACTTTTCTCTGCTCAGTCAGGAGCTG 720  
1214 GAGGCTGACCTTCTCTGAGACCACTCACTGCTATGGAGTGTCTTGGGCTCAGAGC 1273  
721 GAGGCTGACCTTCTCTGAGACCACTCACTGCTATGGAGTGTCTTGGGCTCAGAGC 780  
1274 GAACGTTTCCCATCTAACCTGGAGGTTGAAGCTTCAACACAGCTTCAAGTCAGAGGTA 1333  
781 GAACGTTTCCCATCTAACCTGGAGGTTGAAGCTTCAACACAGCTTCAAGTCAGAGGTA 840  
1334 AATGCTTCTCTCTTTGGAATCTGCCCCATGTGAAAATGAGCCCTCAAGAAAAGTGAAGAA 1393  
841 AATGCTTCTCTCTTTGGAATCTGCCCCATGTGAAAATGAGCCCTCAAGAAAAGTGAAGAA 900  
1394 GGCAATGTCTCTGGCAGTGTGCTGTGGCAGTGTGCTTTCAGAGAGCCTATGTCAGGC 1453  
901 GGCAATGTCTCTGGCAGTGTGCTGTGGCAGTGTGCTTTCAGAGAGCCTATGTCAGGC 960  
1454 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCAC 1513  
961 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGTTCCAC 1020  
1514 TCCACTGCAGCCTCATGGGGTCTCCCTCTTTTCAACACGCGCTGCAAGAAAGAGGATG 1573  
1021 TCCACTGCAGCCTCATGGGGTCTCCCTCTTTTCAACACGCGCTGCAAGAAAGAGGATG 1080  
1574 AGGAAAATATAA 1585  
1081 AGGAAAATATAA 1092

```
US-09-461-697-92
; Sequence 92, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-92

Query Match      35.1%; Score 600; DB 4; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.3e-161;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 986 ATGGAGCAGGTATTCCATGAAGTGGGTATTTGGCAGTGTCTCTCCCTCCAGAAAGTTCTGG 1045
DB 1 ATGGAGCAGGTATTCCATGAAGTGGGTATTTGGCAGTGTCTCTCCCTCCAGAAAGTTCTGG 60

QY 1046 CAGCACCACATCAAGGACTATACAGTTACATGTACAGATTAGTACGCACTCTCTGAA 1105
DB 61 CAGCACCACATCAAGGACTATACAGTTACATGTACAGATTAGTACGCACTCTCTGAA 120

QY 1106 GAATATGAAGGATTTGTAATCTCTGAGAAGGCCACAGAGGACGCTTAACCTGTGAAGATC 1165
DB 121 GAATATGAAGGATTTGTAATCTCTGAGAAGGCCACAGAGGACGCTTAACCTGTGAAGATC 180

QY 1166 AAGGAGAACCTGTGAGGACATACATTTTCTCTGAGTGGAGCTGGAGGCTGACCTT 1225
DB 181 AAGGAGAACCTGTGAGGACATACATTTTCTCTGAGTGGAGCTGGAGGCTGACCTT 240

QY 1226 GCTTCTGAGACCACTGACTCCCTATGGAGTGTCTGGGCTCAGAGCGACGCTTCCA 1285
DB 241 GCTTCTGAGACCACTGACTCCCTATGGAGTGTCTGGGCTCAGAGCGACGCTTCCA 300

QY 1286 TCTAACCTGGAGTTGAAGCTTCCACACAGGCTTCAAGTGAGAGGTTAAATGCTTCTCT 1345
DB 301 TCTAACCTGGAGTTGAAGCTTCCACACAGGCTTCAAGTGAGAGGTTAAATGCTTCTCT 360

QY 1346 CTTTGGAAATCGGCCCATGTGAAAATGGAGCCTCAAGAAAGTGAAGGCAATGTCTCT 1405
DB 361 CTTTGGAAATCGGCCCATGTGAAAATGGAGCCTCAAGAAAGTGAAGGCAATGTCTCT 420

QY 1406 GGGATGTGTGCTGGGCGAGTGTCTTTCAGGAGGCTATGTTCAGGACATGAGTGAAGCT 1465
DB 421 GGGATGTGTGCTGGGCGAGTGTCTTTCAGGAGGCTATGTTCAGGACATGAGTGAAGCT 480

QY 1466 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGGTTCCTCCTCCTCAGCAGC 1525
DB 481 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGGTTCCTCCTCCTCAGCAGC 540

QY 1526 CTCATGGGTCCTCCCTGTTTTCACACAGGCTTCAAGAGAGAGGATGAGGAAATATAA 1585
DB 541 CTCATGGGTCCTCCCTGTTTTCACACAGGCTTCAAGAGAGAGGATGAGGAAATATAA 600

RESULT 5
US-09-461-697-96
; Sequence 96, Application US/09461697
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-96

Query Match      29.8%; Score 510; DB 4; Length 510;
Best Local Similarity 100.0%; Pred. No. 5.9e-136;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1076 ATGCTACAGATTACTAAGCAACTCTCTGAAGAATATGAAGGATTTCAATCTCTGAGAAG 1135
DB 1 ATGCTACAGATTACTAAGCAACTCTCTGAAGAATATGAAGGATTTCAATCTCTGAGAAG 60

QY 1136 GCCACAGAGACGCTAAACCTGTGAAGATCAAGAGGAAACCTGTGAGCGACATCACATTTT 1195
DB 61 GCCACAGAGACGCTAAACCTGTGAAGATCAAGAGGAAACCTGTGAGCGACATCACATTTT 120

QY 1196 CCTGTCTAGTGAGGAGCTGGAGGCTGACCTTCTTCTTGAGACCACTGCTGCTTATGGGA 1255
DB 121 CCTGTCTAGTGAGGAGCTGGAGGCTGACCTTCTTCTTGAGACCACTGCTGCTTATGGGA 180

QY 1256 GTGCTTGGGCTCAGAGCGACGCTTCCCATCTAACCTGAGGTTGAAGCTTCAACACAG 1315
DB 181 GTGCTTGGGCTCAGAGCGACGCTTCCCATCTAACCTGAGGTTGAAGCTTCAACACAG 240

QY 1316 GCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGAATCTGGCCCATGTGAAAATGGAG 1375
DB 241 GCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGAATCTGGCCCATGTGAAAATGGAG 300

QY 1376 CCTCAAGAAAGTGAAGAGGCAATGCTCTGCGGCATGGTGTGTGGGCACTGATGTCTTC 1435
DB 301 CCTCAAGAAAGTGAAGAGGCAATGCTCTGCGGCATGGTGTGTGGGCACTGATGTCTTC 360

QY 1436 GAGGAGCCTATGTCAGGCACTGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGAT 1495
DB 361 GAGGAGCCTATGTCAGGCACTGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGAT 420

QY 1496 AGCAGCTATGTTTCCCACTCCACTGACAGCCTCATGGGGTCTCTCCCTGTTTCAACACAG 1555
DB 421 AGCAGCTATGTTTCCCACTCCACTGACAGCCTCATGGGGTCTCTCCCTGTTTCAACACAG 480

QY 1556 CGCTGCAAGAGAGGATGAGGAAATATAA 1585
DB 481 CGCTGCAAGAGAGGATGAGGAAATATAA 510

RESULT 6
US-09-461-697-100
; Sequence 100, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
```

; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 100  
; LENGTH: 336  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-100

Query Match 19.6%; Score 336; DB 4; Length 336;  
Best Local Similarity 100.0%; Pred. No. 2.2e-86;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1250 ATGGAGTGTGTTGGGCTCAGAGCAACGCTTCCATCTAACCTGAGAGTTGAAGCTTCA 1309  
Db 1 ATGGAGTGTGTTGGGCTCAGAGCAACGCTTCCATCTAACCTGAGAGTTGAAGCTTCA 60  
Qy 1310 CCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGGAACTCTGCGCCCATGTGAAA 1369  
Db 61 CCACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGGAACTCTGCGCCCATGTGAAA 120  
Qy 1370 ATGGAGGCTCAAGAACTGAAGAGCAATGTCTCTGGGATGTTGCTGGGCAATGAT 1429  
Db 121 ATGGAGGCTCAAGAACTGAAGAGCAATGTCTCTGGGATGTTGCTGGGCAATGAT 180  
Qy 1430 GTCTTCAGGAGCTATGTCCAGCATGAGTGAAGCTTCTCTCAGAGCCCTGATGAC 1489  
Db 181 GTCTTCAGGAGCTATGTCCAGCATGAGTGAAGCTTCTCTCAGAGCCCTGATGAC 240  
Qy 1490 TCAGATAGCAGTATGTTCCCACTCCACTCCAGCCTCATGGGCTCTCCCTGTTTC 1549  
Db 241 TCAGATAGCAGTATGTTCCCACTCCACTCCAGCCTCATGGGCTCTCCCTGTTTC 300  
Qy 1550 AACCAGCGCTGCAAGAGAGGATGAGGAAATATAA 1585  
Db 301 AACCAGCGCTGCAAGAGAGGATGAGGAAATATAA 336

RESULT 7  
US-09-461-697-104  
; Sequence 104, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 104  
; LENGTH: 216  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-104

Query Match 12.6%; Score 216; DB 4; Length 216;

Best Local Similarity 100.0%; Pred. No. 3e-52;  
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1370 ATGGAGGCTCAAGAACTGAAGAGCAATGTCTCTGGGATGTTGCTGGGCAATGAT 1429  
Db 1 ATGGAGGCTCAAGAACTGAAGAGCAATGTCTCTGGGATGTTGCTGGGCAATGAT 60  
Qy 1430 GTCTTCAGGAGCTATGTCCAGCATGAGTGAAGCTTCTCTCAGAGCCCTGATGAC 1489  
Db 61 GTCTTCAGGAGCTATGTCCAGCATGAGTGAAGCTTCTCTCAGAGCCCTGATGAC 120  
Qy 1490 TCAGATAGCAGTATGTTCCCACTCCACTCCAGCCTCATGGGCTCTCCCTGTTTC 1549  
Db 121 TCAGATAGCAGTATGTTCCCACTCCACTCCAGCCTCATGGGCTCTCCCTGTTTC 180  
Qy 1550 AACCAGCGCTGCAAGAGAGGATGAGGAAATATAA 1585  
Db 181 AACCAGCGCTGCAAGAGAGGATGAGGAAATATAA 216

RESULT 8  
US-09-461-697-94  
; Sequence 94, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 94  
; LENGTH: 159  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-94

Query Match 9.3%; Score 159; DB 4; Length 159;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1002 ATGAAGTGGTATTTGGCAGTGTCTCTCCCTCCAGAGTTCTGGCAGCAGCCGATCAAG 1061  
Db 1 ATGAAGTGGTATTTGGCAGTGTCTCTCCCTCCAGAGTTCTGGCAGCAGCCGATCAAG 60  
Qy 1062 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTG 1121  
Db 61 ACTATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGAAATATGAAGGATTG 120  
Qy 1122 TCAATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGA 1160  
Db 121 TCAATCCTGAGAGGCCACAGAGGACGCTAAACCTGTGA 159

RESULT 9  
US-09-461-697-112  
; Sequence 112, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.

```
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112
; LENGTH: 141
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-112

Query Match      8.2%; Score 141; DB 4; Length 141;
Best Local Similarity 100.0%; Pred. No. 6.2e-31;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1445 ATGTCAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATGAGCAGCTAT 1504
      |||||||
Db 1 ATGTCAGGCATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATGAGCAGCTAT 60

QY 1505 GGTTCCTCCACTGCATGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCGCTGCAAG 1564
      |||||||
Db 61 GGTTCCTCCACTGCATGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCGCTGCAAG 120

QY 1565 AAGAGGATGAGGAATAATAA 1585
      |||||||
Db 121 AAGAGGATGAGGAATAATAA 141

RESULT 10
US-09-461-697-114
; Sequence 114, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-114

Query Match      7.7%; Score 132; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.2e-28;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1454 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATGAGCAGCTATGTTCCAC 1513
      |||||||
Db 1 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATGAGCAGCTATGTTCCAC 60

QY 1514 TCCACTGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCGCTGCAAGAAGAGGATG 1573
      |||||||
Db 61 TCCACTGACAGCCTCATGGGGTCTCCCTGTTTCAACCGCGCTGCAAGAAGAGGATG 120

QY 1574 AGGAATAATAA 1585
      |||||||
```

```
Db 121 AGGAAATATAA 132

RESULT 11
US-09-461-697-78
; Sequence 78, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 105
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-78

Query Match      6.1%; Score 105; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 1e-20;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 ATGACCCACCCCTGCACCAACCCCTCAGCCACACAGCCGAGCCGCCACTATGCTGGACA 503
      |||||||
Db 1 ATGACCCACCCCTGCACCAACCCCTCAGCCACACAGCCGAGCCGCCACTATGCTGGACA 60

QY 504 TCCCTCTCAGAGCCATGTAGTCTCACCATCCATCAGATTCAAGTTGA 548
      |||||||
Db 61 TCCCTCTCAGAGCCATGTAGTCTCACCATCCATCAGATTCAAGTTGA 105

RESULT 12
US-09-461-697-88
; Sequence 88, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 88
; LENGTH: 96
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-88

Query Match      5.6%; Score 96; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 3.5e-18;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 ATGTGGCACATGAGTATTCCTTAAGTTTACCAAGTTGCTGCTTTTGTGCTGGACCGGG 950
```

Db 1 ATGTGGCACATGAGTATGCTTAAGTTTACCAAGTTGCTGCGCTTTTCTGCTGACCGGG 60  
QY 951 AGGCCCGGCTGGGACAGACTCCTTTTCCCTGATGGA 986  
Db 61 AGGCCCGGCTGGGACAGACTCCTTTTCCCTGATGGA 96

RESULT 13  
US-09-461-697-116  
; Sequence 116, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 116  
; LENGTH: 90  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-116

Query Match 5.3%; Score 90; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.8e-16;  
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1485 ATGACTCAGATAGCAGTATGTTCCACCTCCACTGACAGCCTCATGGGTCCTCCCTG 1544  
Db 1 ATGACTCAGATAGCAGTATGTTCCACCTCCACTGACAGCCTCATGGGTCCTCCCTG 60

QY 1545 TTTTCACACGCGCTGCAAGAGGATGA 1574  
Db 61 TTTTCACACGCGCTGCAAGAGGATGA 90

RESULT 14  
US-09-461-697-90  
; Sequence 90, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 90  
; LENGTH: 87  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-90

Query Match 5.1%; Score 87; DB 4; Length 87;  
Best Local Similarity 100.0%; Pred. No. 1.2e-15;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 ATGAGTATTCCTTAAGTTTACCAAGTTGCTGCTGTTGGACCGGAGCGCGGC 959  
Db 1 ATGAGTATTCCTTAAGTTTACCAAGTTGCTGCTGTTGGACCGGAGCGCGGC 60

QY 960 TGGGACAGACTCCTTTTCCCTGATGGA 986  
Db 61 TGGGACAGACTCCTTTTCCCTGATGGA 87

RESULT 15  
US-09-461-697-118  
; Sequence 118, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 118  
; LENGTH: 72  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-118

Query Match 4.2%; Score 72; DB 4; Length 72;  
Best Local Similarity 100.0%; Pred. No. 2.2e-11;  
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1503 ATGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACACGCGCTGCA 1562  
Db 1 ATGTTCCCACTCCACTGACAGCCTCATGGGTCCTCCCTGTTTCAACACGCGCTGCA 60

QY 1563 AGAAGAGGATGA 1574  
Db 61 AGAAGAGGATGA 72

Search completed: July 18, 2003, 09:48:24  
Job time : 92 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2003, 07:39:02 ; Search time 424 Seconds  
(without alignments)  
9087.663 Million cell updates/sec

Title: US-09-857-308-2  
Perfect score: 1711  
Sequence: 1 acgcgatcttgcctcaggc.....aaaaaaaaaaaaaaaaaaaaa 1711

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1711	100.0	1711	21	Human ART-1 nucleo
2	1676.8	98.0	6026	22	Human protein enco
3	1309	76.5	1468	21	Human secreted pro
4	1288	75.3	1825	22	Human cell death p
5	1233.2	72.1	1239	22	Human cell death p
6	1225.6	71.6	2606	24	Human cDNA for nov
7	1092	63.8	1092	22	Human cell death p
8	850.8	49.7	1046	20	Human endometrium
9	600	35.1	600	22	Human cell death p

10	510	29.8	510	22	AAH84181	Human cell death p
11	468	27.4	544	21	AAC08410	Human secreted pro
12	408	23.8	12244	22	AAK89991	Human digestive sy
13	336	19.6	336	22	AAH84183	Human cell death p
14	293.8	17.2	517	23	ABK41954	CDNA encoding nove
15	216	12.6	216	22	AAH84185	Human cell death p
16	159	9.3	159	22	AAH84180	Human cell death p
17	141	8.2	141	22	AAH84189	Human cell death p
18	132	7.7	132	22	AAH84190	Human cell death p
19	109	6.4	2954	22	AAH14356	Human cDNA sequenc
20	108	6.3	554	22	AAH06243	Human cDNA clone (
21	105	6.1	105	22	AAH84172	Human cell death p
22	96	5.6	96	22	AAH84177	Human cell death p
23	90	5.3	90	22	AAH84191	Human cell death p
24	87	5.1	87	22	AAH84178	Human cell death p
25	72	4.2	72	22	AAH84192	Human cell death p
26	69	4.0	69	22	AAH84175	Human cell death p
27	60	3.5	60	22	AAH84186	Human cell death p
28	57	3.3	57	22	AAH84193	Human cell death p
29	51	3.0	51	22	AAH84182	Human cell death p
30	48.4	2.8	1340	18	AAV24018	Human neuro-D gene
31	48.4	2.8	1844	19	AAV24018	Human BHF1 coding
32	48.4	2.8	2502	21	AAK62679	Human NeuroD1 gene
33	48	2.8	48	22	AAH84187	Human cell death p
34	46.8	2.7	442	22	AAI83061	Human polynucleoti
35	44	2.6	1654	21	AAFI8137	Lung cancer associ
36	43.6	2.5	2089	16	AAI05513	Murine neurogenic
37	43.6	2.5	2089	18	AAV74887	Mouse neurogenic d
38	43.6	2.5	2089	19	AAV42928	CDNA encoding muri
39	42.4	2.5	2572	23	AAH87900	DNA encoding novel
40	42.2	2.5	168	23	ABV57087	Human prostate exp
41	42.2	2.5	1708	22	AAH35032	Human colon cancer
42	42	2.5	520	22	AAK60634	Human cancer agent
43	42	2.5	2329	22	AAV97884	Human secreted pro
44	42	2.5	2355	22	AAV97900	Human secreted pro
45	42	2.5	2500	21	AAV21701	Human breast and o

ALIGNMENTS

RESULT 1  
AAAG62864  
ID AAA62864 standard; DNA: 1711 BP.  
AC  
XX  
AC  
AAAG62864;  
XX  
26-OCT-2000 (first entry)  
DT  
XX  
Human ART-1 nucleotide sequence.  
DE  
XX  
Human; tumour antigen protein; ART-1; HLA antigen; cytostatic; ds;  
KW bound cytotoxic T cell; tumour; cancer; remedy; prevent; diagnose.  
XX  
Homo sapiens.  
OS  
XX  
WO200032770-A1.  
PN  
XX  
08-JUN-2000.  
PD  
XX  
30-NOV-1999; 99WO-JP06682.  
PF  
XX  
01-DEC-1998; 98JP-0341253.  
PR  
XX  
(SUMU ) SUMITOMO PHARM CO LTD.  
PA (ITOH/) ITOH K.  
XX  
Itoh K, Gomi S;  
XX  
WPI: 2000-412318/35.  
DR P-PSDB; AAB03880.  
XX  
Novel tumor antigen protein ART-1, tumor antigen peptide originating

PT from it, their derivatives, and DNAs, applicable in vivo or in vitro as PT remedies, preventives and diagnostics for tumors -



PS Claim 2; Page 46-49; 59pp; Japanese.

The invention relates to a novel human tumour antigen protein, ART-1. Included in the invention are polynucleotide sequences encoding the ART-1 protein, and mutated ART-1 proteins which when broken down intracellularly produce a tumour antigen peptide that can recognise HLA antigen and bound cytotoxic T cells. Antibodies which specifically recognise ART-1 and its derivative peptides, are also included in the invention. ART-1 exhibits cytostatic activity. The tumour antigen protein, tumour antigen peptide originating from it, their derivatives, and DNAs are applicable in vivo or in vitro as remedies, preventives and diagnostics for tumours.

CC The present sequence DNA encoding the human ART-1 protein.

SQ Sequence 1711 BP; 431 A; 452 C; 431 G; 397 T; 0 other;

```
Query Match      100.0%; Score 1711; DB 21; Length 1711;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1711; Conservative 0; Mismatches 0; Indels 0;
```

QY	1	ACGCAGATCCTTGCCCTCAGGCCCTCTCGAGGTCACAGACAGCGCCGACGCCCGCTCTCGCAGC	60
DB	1	ACGCAGATCCTTGCCCTCAGGCCCTCTCGAGGTCACAGACAGCGCCGACGCCCGCTCTCGCAGC	60
QY	61	CAGCAGTGAATAGTGTGGTACTCCTCTTGCTCGGTTCCAGGTCAGAGTCCCGCGTCTTCC	120
DB	61	CAGCAGTGAATAGTGTGGTACTCCTCTTGCTCGGTTCCAGGTCAGAGTCCCGCGTCTTCC	120
QY	121	GGTGGCCCTGAACGTCACGGCAGCTCAGGACCCCTGTGATTGGCGCCCTGGCGCGCGGAC	180
DB	121	GGTGGCCCTGAACGTCACGGCAGCTCAGGACCCCTGTGATTGGCGCCCTGGCGCGCGGAC	180
QY	181	GTGACCGGAGAAACCCCTGGAGGACATTGGGCAATTCCTTGGCTCCGTGCTGTCTTCTCG	240
DB	181	GTGACCGGAGAAACCCCTGGAGGACATTGGGCAATTCCTTGGCTCCGTGCTGTCTTCTCG	240
QY	241	TGCTCCTTTTCGGGCAAGGATCTCAATATTCAGTCTTTGACCGACACAGAAATGCCCTGGCA	300
DB	241	TGCTCCTTTTCGGGCAAGGATCTCAATATTCAGTCTTTGACCGACACAGAAATGCCCTGGCA	300
QY	301	TTTGATAAATGTTTGTGTAACCTTGAGAGACATATGACAAATGAATCTGCAAGAATACTG	360
DB	301	TTTGATAAATGTTTGTGTAACCTTGAGAGACATATGACAAATGAATCTGCAAGAATACTG	360
QY	361	GGGAGAGATACCAATATCATCAAGCCAGACCAACAGAAAGTTCCCTCGATTTGCTCCACG	420
DB	361	GGGAGAGATACCAATATCATCAAGCCAGACCAACAGAAAGTTCCCTCGATTTGCTCCACG	420
QY	421	GGAGTTCGCTCTGCTGGAGTCCATGACCCACCCCTGCACCAACCTCAGCCCAACAGCC	480
DB	421	GGAGTTCGCTCTGCTGGAGTCCATGACCCACCCCTGCACCAACCTCAGCCCAACAGCC	480
QY	481	GAAAGCCCCCACTATGCTGGACATCCCTCAGAGCCATGAGTCTCACATCCATACGAT	540
DB	481	GAAAGCCCCCACTATGCTGGACATCCCTCAGAGCCATGAGTCTCACATCCATACGAT	540
QY	541	TCAGTTGATTACGACAACCCAGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCCAGAA	600
DB	541	TCAGTTGATTACGACAACCCAGCTCTTCGCAACCTTATTGCCACAGCTCAGGCCCAGAA	600
QY	601	TCAGCAGCAGCAGAAAGGTGTAAACATGAAGAGAGTGAACCTCTTTCCCTCGTGCCCTCG	660
DB	601	TCAGCAGCAGCAGAAAGGTGTAAACATGAAGAGAGTGAACCTCTTTCCCTCGTGCCCTCG	660
QY	661	GTACCTCCTCTCCCTGATGACCTCCTTGCCCTTTAGATTGTGAAGATCCCAATGCCACATT	720
DB	661	GTACCTCCTCTCCCTGATGACCTCCTTGCCCTTTAGATTGTGAAGATCCCAATGCCACATT	720
QY	721	CCAGATCCGGCAGCTGACCCAGAGATGACTTTTATCGTGGGAAAGGGAACCTGTGAC	780

RESULT 2  
AAH99786/c

AAH99786/C  
ID AAH99786 standard; CDNA; 6026 BP.

AA  
AC  
AAH99786:

XX XX



16-OCT-2001 (first entry)

Human protein encoding cDNA sequence SEO ID NO:621.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus; anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antiophylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.

**Homo sapiens.**

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

23-DEC-1999; 99US-0471275.

23-DEC-1999; 9903-0471Z/3;  
21-JAN-2000; 2000US-0488725;

25-APR-2000; 2000US-0552317.

(HYSE-) HYSEQ INC.

Tanq YT, Liu C, Drmanac RT:

WPI; 2001-457603/49.

WFI, 2001-437803/  
P-PSDB; AAM25845.

Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

Claim 1; Page 647-648; 1217pp: English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antinaemic; antiaggregant; haemostatic; vulnery; antitumor; osteoprotic; dermatological; anti allergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

Sequence 6026 BP; 1725 A; 1288 C; 1386 G; 1627 T; 0 other;

Query Match

Query Match 98.0%; Score 1676.8; DB 22; Length 6026;  
Best Local Similarity 99.6%; Pred. No. 0;

Best Local Similarity 99.6%; pred. No. 0;  
Matches 1693: Conservative 0: Mismatches

QY 1 ACGCGATCCTTGCCTCAGGCCCTCTCAGAGTCCAGACAGCGGCCAGCCCGCTCTGCGAGC 60

Qv

[illegible]



QY	838	GGCCACGCGGGCTTTGAC	TGCTGCTAATGAGAGTGCTCTGGAGACCCCTAACTGATGTGC	897
Db	504	GGCCACGCGGGCTTTGAC	TGTGCTAATGAGAGTGCTCTGGAGACCCCTAACTGATGTGC	563
QY	898	ACATGAGTATTGCCCTTAAG	TTTACCAAGTTTGCCTGCGTTTTCCTGTGTGGACCGGAGGCCCG	957
Db	564	ACATGAGTATTGCCCTTAAG	TTTACCAAGTTTGCCTGCGTTTTCCTGTGTGGACCGGAGGCCCG	623
QY	958	GCTGGGACAGACTCCCTTTTCC	- - -TGATGTGATGGAGCAGGTATTCATGAAGTGGGTAT	1014
Db	624	GCTGGGACAGACTCCCTTTTCC	TTTCCCTTGAATGGAAGGAGCAGGGATTCATGAAGTGGGTAT	683
QY	1015	TGGCAGTGTGCTCTCCCTCCAGAA	GTTCCTGGCAGCACCGCATCAAGGACTATCACAGTTA	1074
Db	684	TGGCAGTGTGCTCTCCCTCCAGAA	GTTCCTGGCAGCACCGCATCAAGGACTATCACAGTTA	743
QY	1075	CATGCTACAGATTAGTAAGCA	ACTCTCTGAAGAATATGAAAGGATTGTCAATCCTGAGAA	1134
Db	744	CATGCTACAGATTAGTAAGCA	ACTCTCTGAAGAATATGAAAGGATTGTCAATCCTGAGAA	803
QY	1135	GGCCACAGAGGACGCTAAAC	CTGTGAAGATCAAGGAGAACCTGTGAGCGACATCACTTT	1194
Db	804	GGCCACAGAGGACGCTAAAC	CTGTGAAGATCAAGGAGAACCTGTGAGCGACATCACTTT	863
QY	1195	TCCTGTCAGTGAAGGCTGAGG	CTGACCTTGCTTCTGGAGACCACTGCTGCTATGGG	1254
Db	864	TCCTGTCAGTGAAGGCTGAGG	CTGACCTTGCTTCTGGAGACCACTGCTGCTATGGG	923
QY	1255	AGTGCTTGGGGCTCAGAGCA	AGCGTTCCCATCTAACCTGGAGGTTGAAGCTTCACCACA	1314
Db	924	AGTGCTTGGGGCTCAGAGCA	AGCGTTCCCATCTAACCTGGAGGTTGAAGCTTCACCACA	983
QY	1315	GGCTTCAAGTGCAGAGGTAA	ATGCTTCTCCTCTTTGGAATCTGGCCCATGTGAAATGGA	1374
Db	984	GGCTTCAAGTGCAGAGGTAA	ATGCTTCTCCTCTTTGGAATCTGGCCCATGTGAAATGGA	1043
QY	1375	GCCTCAAGAAGTGAAGAAGG	CAATGCTCTGGGCATGGTGCTGGGCACTGATGCTTT	1434
Db	1044	GCCTCAAGAAGTGAAGAAGG	CAATGCTCTGGGCATGGTGCTGGGCACTGATGCTTT	1103
QY	1435	CGAGGAGCCTTATGTCAGG	CATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGA	1494
Db	1104	CGAGGAGCCTTATGTCAGG	CATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGA	1163
QY	1495	TAGCAGTATGTTGCCACTP	CCACTGACAGCCTCATGGGGTCTCCCTGTTTTCACCA	1554
Db	1164	TAGCAGTATGTTGCCACTP	CCACTGACAGCCTCATGGGGTCTCCCTGTTTTCACCA	1223
QY	1555	GCCTCAGAAGCAGATGAGG	AAATATATAAAGGAAAGAGGGAGATGTTTGTCCAGA	1614
Db	1224	GCCTCAGAAGCAGATGAGG	AAATATATAAAGGAAAGAGGGAGATGTTTGTCCAGA	1283
QY	1615	CCTACTAGACCCACAGAA	AGGTTTTTGTATTAGAAATCTGTTTCCTTAAAAATTCATTT	1674
Db	1284	CCTACTAGACCCACAGAA	AGGTTTTTGTATTAGAAATCTGTTTCCTTAAAAATTCATTT	1343
QY	1675	GACTCCTGTTCTTAAAAA	AAAA 1695	
Db	1344	GACTCCTGTTCTTAAAAA	CAAA 1364	

## RESULT 4

RESULT 4  
AAH84170

AAH84170  
ID AAH84170 standard: cDNA: 1825 BP.

XX  
ID  
AAH84170XX  
AC  
AAH84170:AC  
XX  
AAH84170;

XX 21-SEP-2001 (first entry)

DT 21-SEP-2001 (first entry)  
XX

Human cell death protective cDNA clone CNT-00714 SEQ.75

DE Human cell death protective cDNA clone CNI-00714, SEQ:75.  
XX  
XX

KW	cell death-associated disorder; central nervous system disorder;
KW	psychiatric disorder; neurological disorder; ischemia-related disorder;
KW	stroke; cerebral infarction; ischaemic encephalopathy;
KW	neurodegenerative disorder; Alzheimer's disease; Huntington's disease;
KW	Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;
KW	vascular disease; ophthalmological disorder; diabetic retinopathy;
KW	macular degeneration; hypertension; myocardial infarction;
KW	atherosclerosis; respiratory disorder; asthma; transgenic animal;
KW	chronic obstructive pulmonary disease; neoplastic condition; cancer;
KW	benign tumour; anaemia; gastrointestinal disorder; gastritis;
KW	ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;
KW	glomerulonephritis; cystitis; endometriosis; endocrine disorder;
KW	Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;
KW	urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS; ss
OS	Homo sapiens.

**Homo sapiens.**

XX  
PN  
WO200145638-A2

28-TIN-2001

[illegible]

XX

XX  
XX

FA  
(COGE-) COGENI NEUROSCIENCE INC.  
XX

PI Lo DC, Barney S, T  
XX

XX  
DR WPT: 2001-390297/A1

DR	WPI; 2001-390297/41.
DR	P-PSDB: AAC98644 AAC98645 AAC98646 AAC98647 AAC98648 AAC98649

DR P-PSDB; AAG98644, AAG98645, AAG98646, AAG98647, AAG98648, AAG98649, AAG98650 AAG98651 AAG98652 AAG98653 AAG98654 AAG98655 AAG98656

DR AAG98650, AAG98651, AAG98652, AAG98653, AAG98654, AAG98655, AAG98656,  
DB AAG98657 AAG98658 AAG98659 AAG98660 AAG98661 AAG98662 AAG98663

DR AAG98657, AAG98658, AAG98659,  
DB AAG98660, AAG98661, AAG98662

DR AAG98671, AAG98672, AAG98673.

XX

PT Novel protective sequence polynucleotides and polypeptides, used to

PT identify modulators of their expression and activity, which are used in

PT to treat central nervous system conditions, diseases and disorders -

Example: Fig 1c; 325pp; English.

Sequences AAH84132-AAH84370 represent human nucleic acid sequences which protect against cell death (i.e., apoptosis or necrosis). Sequences AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265, AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones, while the remaining nucleic acid sequences within the range given above represent the open reading frames (ORFs) of these cDNA clones. Sequences AAG98610-AAG98829 represent the polypeptides encoded by the cell death protective ORFs. The cell death protective cDNA clones are able to prevent, delay or reverse progression through the apoptotic or necrotic pathways when injected into a cell predisposed to or undergoing cell death. The cell death protective nucleic acids and polypeptides can be used in the diagnosis and treatment of disorders associated with cell death, and to screen for compounds which modulate their activity or expression. Such modulators, preferably a small organic molecule, an antibody, a ribozyme, or an antisense molecule, can also be used to treat cell death-related diseases. Such diseases include those associated with the central nervous system including psychiatric or neurological disorders, especially ischaemia-related conditions such as strokes, and also includes neurodegenerative disorders such as Alzheimer's disease, Huntington's disease, or Parkinson's disease. The modulators may also be used to treat infections such as meningitis, malaria, or trypanosomiasis; vascular diseases such as ischaemic encephalopathy or cerebral infarction; eye conditions such as diabetic retinopathy or macular degeneration; hypertension; myocardial infarction; atherosclerosis; respiratory conditions such as asthma or chronic obstructive pulmonary disease; neoplastic conditions such as cancers or benign tumours; blood cell conditions such as anaemia; gastrointestinal conditions such as gastritis or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney disorders such as glomerulonephritis; cystitis; endometriosis; endocrine disorders such as Grave's disease or Hashimoto's thyroiditis; skin conditions such as dermatitis or urticaria; or immune





KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;  
 KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;  
 KW lymphoid cell deficiency.  
 XX Homo sapiens.  
 XX  
 PN WO200177290-A2.  
 PD 18-OCT-2001.  
 XX  
 XX 29-MAR-2001; 2001WO-US10295.  
 PF  
 XX 06-APR-2000; 2000US-194941P.  
 PR  
 XX (GEMY ) GENETICS INST INC.  
 PA  
 XX  
 XX  
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 PI Gulukota K, Graham JR;  
 XX  
 DR WPI; 2002-179323/23.  
 XX  
 XX Six hundred and twenty five polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders  
 PT .  
 XX  
 PS Claim 1; Page 315-316; 339pp; English.  
 XX  
 CC The invention relates to 625 polynucleotides which have been derived from  
 CC a variety of human tissue sources and which encode novel secreted  
 CC proteins, their complements and sequences that hybridize to them.  
 CC Also included are a vector comprising the polynucleotide, a host cell  
 CC transformed with the vector, the proteins encoded by the  
 CC polynucleotides, antibodies that bind to the proteins and identification  
 CC of modulators of the proteins or the expression of the polynucleotide.  
 CC The polynucleotides can be used as probes for the identification  
 CC and isolation of full length cDNA and genomic DNA. The polynucleotides  
 CC and proteins can also be used as nutritional supplements. The protein  
 CC is useful in the treatment of various immune deficiencies and disorders  
 CC such as viral infections, bacterial infections, fungal infections,  
 CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,  
 CC autoimmune thyroiditis and diabetes) and allergic reactions and  
 CC conditions (e.g. asthma). They are also useful for treating  
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
 CC useful for tissue regeneration, for wound healing and in the treatment  
 CC of burns, incisions and ulcers. The proteins are also useful for  
 CC regulating haematopoiesis, for treating myeloid or lymphoid cell  
 CC deficiencies. The present sequence is one of the 625 cDNA sequences  
 CC encoding a secreted protein.  
 XX  
 SQ Sequence 2606 BP; 697 A; 612 C; 529 G; 768 T; 0 other;  
 Query Match 71.6%; Score 1225.6; DB 24; Length 2606;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1228; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 464 CCTCAGCCCAACGCGAAGCCCGCCCACTATGCTGGACATCCCTCAGAGCCATGTAGT 523  
 DB 1 CCTCAGCCCAACGCGAAGCCCGCCCACTATGCTGGACATCCCTCAGAGCCATGTAGT 60  
 QY 524 CTCACCATCCATACGATTTCAGTTGATTCAGCACACACCGACGCTCTTCGCAACCTTATTGCC 583  
 DB 61 CTCACCATCCATACGATTTCAGTTGATTCAGCACACACCGACGCTCTTCGCAACCTTATTGCC 120  
 QY 584 ACAGCTCAGGCCGAGCAATCAGACAGACAGAGGTGTAAACACTGAAGAGAGTGAACCT 643  
 DB 121 ACAGCTCAGGCCGAGCAATCAGACAGACAGAGGTGTAAACACTGAAGAGAGTGAACCT 180  
 QY 644 CTTCCTCGTCCCTGGGTGACCTCTCTCCCTGATGACCTCTCCCTTTAGATTGTAAAG 703  
 DB 181 CTTCCTCGTCCCTGGGTGACCTCTCTCCCTGATGACCTCTCCCTTTAGATTGTAAAG 240

RESULT 7

AAH84173

ID AAH84173 standard; cDNA; 1092 BP.

QY 704 AATCCCAATGCACCATTCAGATCCGCGACACAGTACCAGAGAGTGACTTTTATCTGTGG 763  
 DB 241 AATCCCAATGCACCATTCAGATCCGCGACACAGTACCAGAGAGTGACTTTTATCTGTGG 300  
 QY 764 AAAGGGAACTGTGACTGAACCTAGCTGGGCACTCCTGTGCGCAGCTCCTCTACAGGCA 823  
 DB 301 AAAGGGAACTGTGACTGAACCTAGCTGGGCACTCCTGTGCGCAGCTCCTCTACAGGCA 360  
 QY 824 GTGGCCACAATCCTGGCCACGCGGGCTTTGACTCTGCTAAATGACAGAGTGCTCTGAGACC 883  
 DB 361 GTGGCCACAATCCTGGCCACGCGGGCTTTGACTCTGCTAAATGACAGAGTGCTCTGAGACC 420  
 QY 884 CTAAGTGTGGGACACATGAGTATTGCTTAAGTTTACCAAGTTCTCTCGTTTCTGTG 943  
 DB 421 CTAAGTGTGGGACACATGAGTATTGCTTAAGTTTACCAAGTTCTCTCGTTTCTGTG 480  
 QY 944 GACCGGGAGCCCGGCTGGGACAGACTCCTTTTCTGATGTAGTGAGGACAGATATTCAT 1003  
 DB 481 GACCGGGAGCCCGGCTGGGACAGACTCCTTTTCTGATGTAGTGAGGACAGATATTCAT 540  
 QY 1004 GAAGTGGGTATGGCAGTGTCTCTCCCTCCAGAGTTCTGGCAGCACCAGCATCAAGAC 1063  
 DB 541 GAAGTGGGTATGGCAGTGTCTCTCCCTCCAGAGTTCTGGCAGCACCAGCATCAAGAC 600  
 QY 1064 TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGATATATGAAGGATTGTC 1123  
 DB 601 TATCACAGTTACATGCTACAGATTAGTAAGCAACTCTCTGAAGATATATGAAGGATTGTC 660  
 QY 1124 AATCCTGAGAAGGCCACAGAGAGCGCTAAACCTGTGAAGATCAAGAGAGAACTGTGAGC 1183  
 DB 661 AATCCTGAGAAGGCCACAGAGAGCGCTAAACCTGTGAAGATCAAGAGAGAACTGTGAGC 720  
 QY 1184 GACATCACTTTTCTCTGTCAGTGAGGAGCTGGAGGCTGACCTTGTCTTCTGGAGACAGTCA 1243  
 DB 721 GACATCACTTTTCTCTGTCAGTGAGGAGCTGGAGGCTGACCTTGTCTTCTGGAGACAGTCA 780  
 QY 1244 CTGCTATGGAGTGTCTTGGGGCTCAGAGCAAGCGCTTCCCATCTAACCTGAGAGTTGAA 1303  
 DB 781 CTGCTATGGAGTGTCTTGGGGCTCAGAGCAAGCGCTTCCCATCTAACCTGAGAGTTGAA 840  
 QY 1304 GCTTCACACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTCTTGGAACTGGCCCAT 1363  
 DB 841 GCTTCACACAGGCTTCAAGTGCAGAGGTAATGCTTCTCTCTTCTTGGAACTGGCCCAT 900  
 QY 1364 GTGAAAATGGAGCCTCAAGAAGTCAAGAAGCAATGCTCTGGGCATGGTGTCTGGGC 1423  
 DB 901 GTGAAAATGGAGCCTCAAGAAGTCAAGAAGCAATGCTCTGGGCATGGTGTCTGGGC 960  
 QY 1424 AGTGATGTCTTCTGAGGAGCCTTATGTGAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCT 1483  
 DB 961 AGTGATGTCTTCTGAGGAGCCTTATGTGAGGCAATGAGTGAAGCTGGGATTCCTCAGAGCCCT 1020  
 QY 1484 GATGACTCAGATAGAGCTATGTTCCCACTCCACTGACAGCTCATGGGTCTCCCTCCT 1543  
 DB 1021 GATGACTCAGATAGAGCTATGTTCCCACTCCACTGACAGCTCATGGGTCTCCCTCCT 1080  
 QY 1544 GTTTTCAACACGCGCTCAAGAAGAGGATGAGGAAATATATAAGGAAAGAGGAGATG 1603  
 DB 1081 GTTTTCAACACGCGCTCAAGAAGAGGATGAGGAAATATATAAGGAAAGAGGAGATG 1140  
 QY 1604 TTTTGTCCAGACCTACTAGACCCCAACAGAAAAGGTTTTTGTATTAGATCTGTTTCCCTTA 1663  
 DB 1141 TTTTGTCCAGACCTACTAGACCCCAACAGAAAAGGTTTTTGTATTAGATCTGTTTCCCTTA 1200  
 QY 1664 AAAATTGATTTGACTCTCTGTTCTTTAAAAAAA 1695  
 DB 1201 AAAATTGATTTGACTCTCTGTTCTTTAAACACAA 1232



cell conditions such as anaemia; gastrointestinal conditions such as gastritis or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney disorders such as glomerulonephritis; cystitis; endometriosis; endocrine disorders such as Grave's disease or Hashimoto's thyroiditis; skin conditions such as dermatitis or urticaria; or immune system disorders such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may additionally be used to generate animal models of cell death-associated disorders. The present sequence represents a cell death protective ORF.

Sequence 1092 BP; 269 A; 280 C; 287 G; 256 T; 0 Other;

Query Match 63.8%; Score 1092; DB 22; Length 1092;  
Best Local Similarity 100.0%; Pred. No. 5.3e-293;  
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 ATGCTGGACATCCCCTCAGAGCCATGTAGTCTCACCATTCCATCCAGTTCAGTTGATTCAG 553  
DB 1 ATGCTGGACATCCCCTCAGAGCCATGTAGTCTCACCATTCCATCCAGTTCAGTTGATTCAG 60  
QY 554 CACAACCCGAGCTTTCGCAACCTTATTGCCAGCTCAGGCCCAAGTAATCAGCAGACACA 613  
DB 61 CACAACCCGAGCTTTCGCAACCTTATTGCCAGCTCAGGCCCAAGTAATCAGCAGACACA 120  
QY 614 GAAGTGTAAACTGAAGAGAGTGAACCTCTTCCTCTGTCGCTGGGTCACTCCCTC 673  
DB 121 GAAGTGTAAACTGAAGAGAGTGAACCTCTTCCTCTGTCGCTGGGTCACTCCCTC 180  
QY 674 CCGTATGACCTCCTGCCTTTTACATTTGTAAGATCCCAATGCACCATCCAGATCCGGCAC 733  
DB 181 CCGTATGACCTCCTGCCTTTTACATTTGTAAGATCCCAATGCACCATCCAGATCCGGCAC 240  
QY 734 AGTGACCCAGAGAGTGAACCTTTATCGTGGAAAGGGGAACCTGTGACTGAACCTCAGCTGG 793  
DB 241 AGTGACCCAGAGAGTGAACCTTTATCGTGGAAAGGGGAACCTGTGACTGAACCTCAGCTGG 300  
QY 794 CACTCTGTGCGCAGCTCCTTACCAGGAGTGGGCCAACAATCTTGGCCACGCGGGCTTT 853  
DB 301 CACTCTGTGCGCAGCTCCTTACCAGGAGTGGGCCAACAATCTTGGCCACGCGGGCTTT 360  
QY 854 GACTGTGCTAATGAGAGTGTCTTGAGAGCCCTTAAGTGTGGACATGAGTATTTGCTT 913  
DB 361 GACTGTGCTAATGAGAGTGTCTTGAGAGCCCTTAAGTGTGGACATGAGTATTTGCTT 420  
QY 914 AAGTTTACCAGTGTGCTGCTTTTCTGTGGACGGGAGGCCGCTGGGACACACTCCT 973  
DB 421 AAGTTTACCAGTGTGCTGCTTTTCTGTGGACGGGAGGCCGCTGGGACACACTCCT 480  
QY 974 TTTCTGTATGTGATGGAGCAGGTATTTCCATGAAGTGGGTATTTGGCAGTGTCTTCCCTC 1033  
DB 481 TTTCTGTATGTGATGGAGCAGGTATTTCCATGAAGTGGGTATTTGGCAGTGTCTTCCCTC 540  
QY 1034 CAGAAGTCTGCGCACCAGCATCAAGCACTATCACAGTTACATGCTTACAGATTAGTAAG 1093  
DB 541 CAGAAGTCTGCGCACCAGCATCAAGCACTATCACAGTTACATGCTTACAGATTAGTAAG 600  
QY 1094 CAACTCTCGAAGAATATGAAGAGTTCCTCAATCTGAGAAGGCCACAGAGGAGCGCTAAA 1153  
DB 601 CAACTCTCGAAGAATATGAAGAGTTCCTCAATCTGAGAAGGCCACAGAGGAGCGCTAAA 660  
QY 1154 CCGTGAAGATCAAGAGAGAACCTGTGAGCAGACATCACTTTTCTGTGATGAGGAGCTG 1213  
DB 661 CCGTGAAGATCAAGAGAGAACCTGTGAGCAGACATCACTTTTCTGTGATGAGGAGCTG 720  
QY 1214 GAGGCTGACCTTGTCTTGGAGACCACTGCTTATGGAGTGTCTTGGGGCTCAGAGC 1273  
DB 721 GAGGCTGACCTTGTCTTGGAGACCACTGCTTATGGAGTGTCTTGGGGCTCAGAGC 780  
QY 1274 GAACGCTTCCCATCTAACTCGAGGTTCGAAGCTTTCACACAGGCTTCAAGTCGAGAGGTA 1333  
DB 781 GAACGCTTCCCATCTAACTCGAGGTTCGAAGCTTTCACACAGGCTTCAAGTCGAGAGGTA 840  
QY 1334 AATGCTTCTCTTCTTGAATCTTGCCCATCTGAAATATGGAGCTTCAAGAAATGGAAGAA 1393

Db 841 AATGCTTCTCTTGGGAATCTGGCCATCTGAAATGGAGCCTCAAGAAAGTGAAGAA 900  
QY 1394 GGCAATGCTCTGGGCATGGTGTCTGGGAGTGTCTTCGAGGAGCCTATGTGCAGGC 1453  
Db 901 GGCAATGCTCTGGGCATGGTGTCTGGGAGTGTCTTCGAGGAGCCTATGTGCAGGC 960  
QY 1454 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGGTTCCAC 1513  
Db 961 ATGAGTGAAGCTGGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGTATGGTTCCAC 1020  
QY 1514 TCCACTGACAGCCTCATGGGTCTCTCCCTGTCTTTCAACCCAGCGCTGCAAGAAGAGATG 1573  
Db 1021 TCCACTGACAGCCTCATGGGTCTCTCCCTGTCTTTCAACCCAGCGCTGCAAGAAGAGATG 1080  
QY 1574 AGGAATAATATA 1585  
Db 1081 AGGAATAATATA 1092

RESULT 8  
AAZ41981  
ID AAZ41981 standard; cDNA: 1046 BP.  
XX AAZ41981;  
AC AAZ41981;  
XX  
DT 31-JAN-2000 (first entry)  
XX Human endometrium tumour cDNA derived EST 1.  
DE Human endometrium tumour cDNA derived EST 1.  
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
KW treatment; uterine; gene therapy; expressed sequence tag; ss.  
XX Homo sapiens.  
XX DE19817948-A1.  
PN 21-OCT-1999.  
XX 17-APR-1998; 98DE-1017948.  
XX 17-APR-1998; 98DE-1017948.  
PR (META-) METAGEN GES GENOMFORSCHUNG MBH.  
PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
PI WPI: 1999-591957/51.  
XX P-PSDB; AAY59941, AAY59942, AAY59943.  
DR New nucleic acid sequences expressed in uterine cancer tissues, and  
XX derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents - - -

PS Claim 3: Page 164; 444pp: German.  
XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison  
CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AAZ41981-242121 represent  
CC EST fragments derived from a human endometrium tumour cDNA library which  
CC encode the protein sequences represented in AAY59941-y60328.

SQ Sequence 1046 BP; 240 A; 304 C; 262 G; 240 T; 0 other;  
Query Match 49.7%; Score 850.8; DB 20; Length 1046;  
Best Local Similarity 96.1%; Pred. No. 5.9e-226;  
Matches 949; Conservative 0; Mismatches 27; Indels 12; Gaps 7;  
QY 1 ACGCGATCCTTGGCTCAGGCTCTCGAGGTTCAGACAGCCGCCAGCCGCTCTCGCAGC 60  
Db 46 ACGCGATCCTTGGCTCAGGCTCTCGAGGTTCAGACAGCCGCCAGCCGCTCTCGCAGC 105  
QY 61 CAGCAGTCAATAGTGTGTACTCTCTCTCGGTTCAGGTCCAGACCTCCCGCTCTCC 120  
Db 106 CAGCAGTCAATAGTGTGTACTCTCTCTCGGTTCAGGTCCAGACCTCCCGCTCTCC 165  
QY 121 GGCTGCCCTGAACGTCAGGCGACCTTCAGAGCCCTGTGATTTGGCGCTCGCGCGCGAC 180  
Db 166 GGCTGCCCTGAACGTCAGGCGACCTTCAGAGCCCTGTGATTTGGCGCTCGCGCGCGAC 225  
QY 181 GTGACCGAGGAAACCCCTGGAGGACTTGGGCAATTCCTTTGGCTCCGCTGCTTCTTCG 240  
Db 226 GTGACCGAGGAAACCCCTGGAGGACTTGGGCAATTCCTTTGGCTCCGCTGCTTCTTCG 285  
QY 241 TGCTCTCTTC-GGGCAAGGATCTCATTATCAGTCTTTGACCGGACACAGATGCTCGC 299  
Db 286 TGCTCTCTTCGGGGCAAGGATCTCATTATCAGTCTTTGACCGGACACAGATGCTCGC 345  
QY 300 ATTTGATAAATGTTTGAACCTTGAAGAGACATATGACAAATGAATCTGCAAGATACT 359  
Db 346 ATTTGATAAATGTTTGAACCTTGAAGAGACATATGACAAATGAATCTGCAAGATACT 405  
QY 360 GGGGAGAGATACCAATATCATCAAGCCAGACCAAGTTCCTTCGATTTGCTCCAC 419  
Db 406 GGGGAGAGATACCAATATCATCAAGCCAGACCAAGTTCCTTCGATTTGCTCCAC 465  
QY 420 GGGAGTTCCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCACAAGC 479  
Db 466 GGGAGTTCCTGCTGGTGAAGTCCATGACCCACCCCTGCACCAACCCCTCAGCCACAAGC 525  
QY 480 CGAAGCCCCCACTATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGA 539  
Db 526 CGAAGCCCCCACTATGCTGGACATCCCTCAGAGCCATGTAGTCTCACCATCCATACGA 585  
QY 540 TTCAGTTGATTCAGCACAAACCGAGCTTCGCAACCTTATTTGCCACAGCTCAGGCCCAGA 599  
Db 586 TTCAGTTGATTCAGCACAAACCGAGCTTCGCAACCTTATTTGCCACAGCTCAGGCCCAGA 645  
QY 600 ATCAGCAGCAGACAGAGGTCTAAAACCTGAAGAGTGAACCTCTTCCCTCGTCCCTCG 659  
Db 646 ATCAGCAGCAGACAGAGGTCTAAAACCTGAAGAGTGAACCTCTTCCCTCGTCCCTCG 705  
QY 660 GGTCACTCTCTCTCCCTGATGACCTCTCTGCTGCTTTAGATTGTAAGAAATCCCAATGCA 719  
Db 706 GGTCACTCTCTCTCCCTGATGACCTCTCTGCTGCTTTAGATTGTAAGAAATCCCAATGCA 765  
QY 720 TCCAGATCCGCGACAGTGAACCCAGAGAGTGAACCTTTATCGTGGGAAAGGGAACCTGTA 779  
Db 766 TCCAGATCCGCGACAGTGAACCCAGAGAGTGAACCTTTATCGTGGGAAAGGGAACCTGTA 825  
QY 780 CTGAAGTCAAGTGGCAGCTCTGTCGGCAGCTCTCTACCC-AGGCAGTGGCCACATCTCG 838  
Db 826 CTGAAGTCAAGTGGCAGCTCTGTCGGCAGCTCTCTACCCAGGAGGAGGAGGAGGAGGAG 885  
QY 839 GCCCAGC--CGGGCTTTGACTGTCTGAATAGAGTGTCTCTGG-AGACCTTAAT--GATG 893  
Db 886 GCCAAGCGCGGCTTTGACTGTCTGAATAGAGTGTCTCTGGAAGACCTTAATGATGT 945  
QY 894 TGGCAGATGATG--CCTTAAGTTTACCAAGTGTGCTGCG---TTTTGCTGTGAGCCG 948  
Db 946 TGGCAGATGATGTTGGCTTTAAAGTTTACCAAGTGTGCTGCTGCTGCTGCTGCTGAGCG 1005  
QY 949 GGAGCCCGGCTGGGACAGACTCCCTTT 976  
Db 1006 GGAAGCCCGGCTGGGAGAGACTTCCTTT 1033



CC vascular diseases such as ischaemic encephalopathy or cerebral  
CC infarction; eye conditions such as diabetic retinopathy or macular  
CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
CC respiratory conditions such as asthma or chronic obstructive pulmonary  
CC disease; neoplastic conditions such as cancers or benign tumours; blood  
CC cell conditions such as anaemia; gastrointestinal conditions such as  
CC gastritis or ulcerative colitis; liver conditions such as biliary  
CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
CC nucleic acids may additionally be used to generate animal models of  
CC cell death-associated disorders. The present sequence represents a  
XX cell death protective ORF.  
SQ Sequence 600 BP; 157 A; 138 C; 166 G; 139 T; 0 other;  
Query Match 35.1%; Score 600; DB 22; Length 600;  
Best Local Similarity 100.0%; Pred. No. 2.3e-156;  
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 986 ATGAGCAGGATATCCATGAAGTGGTATTTGGCAGTGTGCTCCCTCCAGAGTCTGG 1045  
Db 1 ATGAGCAGGATATCCATGAAGTGGTATTTGGCAGTGTGCTCCCTCCAGAGTCTGG 60  
Qy 1046 CAGCACCAGTCAAGGACTATCACAGTTTACATGCTACAGATTAGTAAAGCAACTCTCGAA 1105  
Db 61 CAGCACCAGTCAAGGACTATCACAGTTTACATGCTACAGATTAGTAAAGCAACTCTCGAA 120  
Qy 1106 GAATATGAAGGATTTGCAATCTGAGAAGGCCACAGAGCGCTAAACCTGTGAAGATC 1165  
Db 121 GAATATGAAGGATTTGCAATCTGAGAAGGCCACAGAGCGCTAAACCTGTGAAGATC 180  
Qy 1166 AAGGAGAACCTGTGAGGACATCATTCTCTCAGTGGAGGCTGGAGGCTGACCTT 1225  
Db 181 AAGGAGAACCTGTGAGGACATCATTCTCTCAGTGGAGGCTGGAGGCTGACCTT 240  
Qy 1226 GCTTCTGAGACCACTGCTGCTTATGGAGTGTCTGGGGCTCAGAGCGAAGCTTCCCA 1285  
Db 241 GCTTCTGAGACCACTGCTGCTTATGGAGTGTCTGGGGCTCAGAGCGAAGCTTCCCA 300  
Qy 1286 TCTAACCTGGAGGTTGAAGCTTCAACAGAGCTTCAAGTGCAGAGGTAATGCTTCCT 1345  
Db 301 TCTAACCTGGAGGTTGAAGCTTCAACAGAGCTTCAAGTGCAGAGGTAATGCTTCCT 360  
Qy 1346 CTTTGGATCTGCCCATGTGAATGAGCTCAAGAAATGAGGCTCAAGAAATGAGGCTCAATGCTCT 1405  
Db 361 CTTTGGATCTGCCCATGTGAATGAGCTCAAGAAATGAGGCTCAAGAAATGAGGCTCAATGCTCT 420  
Qy 1406 GGGCATGCTGTGCTGGGCACTGATGCTTTCAGAGGCTTATGTCAGGCACTGAGTGAAGCT 1465  
Db 421 GGGCATGCTGTGCTGGGCACTGATGCTTTCAGAGGCTTATGTCAGGCACTGAGTGAAGCT 480  
Qy 1466 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGGTTCCTCCACTCCAGCAGC 1525  
Db 481 GGGATTCCTCAGAGCCCTGATGACTCAGATAGCAGCTATGGTTCCTCCACTCCAGCAGC 540  
Qy 1526 CTGATGGGCTCTCCCTCTGTTTCAACAGGCTGCAAGAGAGGATGAGGAAATATAA 1585  
Db 541 CTGATGGGCTCTCCCTCTGTTTCAACAGGCTGCAAGAGAGGATGAGGAAATATAA 600  
RESULT 10  
AAH84179  
ID AAH84179 standard; cDNA; 600 BP.  
XX  
AC AAH84179;  
XX  
DT 21-SEP-2001 (first entry)  
XX  
DE Human cell death protective cDNA clone CNI-00714 ORF9, SEQ:92.  
XX  
KW Cell death protective; apoptosis; necrosis; human; drug screening;  
KW cell death-associated disorder; central nervous system disorder;  
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
KW stroke; cerebral infarction; ischaemic encephalopathy;  
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
KW benign tumour; anaemia; gastrointestinal disorder; gastritis; disorder;  
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;  
KW open reading frame; ORF; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200145638-A2.  
XX  
PD 28-JUN-2001.  
XX  
PE 11-DEC-2000; 2000WO-US33547.  
XX  
PF 14-DEC-1999; 99US-0461697.  
XX  
PR (COGE-) COSENT NEUROSCIENCE INC.  
XX  
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
XX  
DR WPI; 2001-390297/41.  
DR P-PSDB; AAG98652.  
XX  
PT Novel protective sequence polynucleotides and polypeptides, used to  
PT identify modulators of their expression and activity, which are used in  
PT to treat central nervous system conditions, diseases and disorders -  
XX  
PS Claim 2; Fig 6I; 325pp; English.  
XX  
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
CC while the remaining nucleic acid sequences within the range given above  
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
CC protective ORFs. The cell death protective cDNA clones are able to  
CC prevent, delay or reverse progression through the apoptotic or necrotic  
CC pathways when injected into a cell predisposed to or undergoing cell  
CC death. The cell death protective nucleic acids and polypeptides can be  
CC used in the diagnosis and treatment of disorders associated with cell  
CC death, and to screen for compounds which modulate their activity or  
CC expression. Such modulators, preferably a small organic molecule, an  
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
CC cell death-related diseases. Such diseases include those associated with  
CC the central nervous system including psychiatric or neurological  
CC disorders, especially ischaemia-related conditions such as strokes, and  
CC also includes neurodegenerative disorders such as Alzheimer's disease,  
CC Huntington's disease, or Parkinson's disease. The modulators may also be  
CC used to treat infections such as meningitis, malaria, or trypanosomiasis;

AAH84179  
ID AAH84179 standard; cDNA; 600 BP.  
XX  
AC AAH84179;  
XX  
DT 21-SEP-2001 (first entry)  
XX  
DE Human cell death protective cDNA clone CNI-00714 ORF9, SEQ:92.  
XX  
KW Cell death protective; apoptosis; necrosis; human; drug screening;  
KW cell death-associated disorder; central nervous system disorder;  
KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
KW stroke; cerebral infarction; ischaemic encephalopathy;  
KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
KW benign tumour; anaemia; gastrointestinal disorder; gastritis; disorder;  
KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;  
KW open reading frame; ORF; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200145638-A2.  
XX  
PD 28-JUN-2001.  
XX  
PE 11-DEC-2000; 2000WO-US33547.  
XX  
PF 14-DEC-1999; 99US-0461697.  
XX  
PR (COGE-) COSENT NEUROSCIENCE INC.  
XX  
PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
XX  
DR WPI; 2001-390297/41.  
DR P-PSDB; AAG98652.  
XX  
PT Novel protective sequence polynucleotides and polypeptides, used to  
PT identify modulators of their expression and activity, which are used in  
PT to treat central nervous system conditions, diseases and disorders -  
XX  
PS Claim 2; Fig 6I; 325pp; English.  
XX  
CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
CC while the remaining nucleic acid sequences within the range given above  
CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
CC protective ORFs. The cell death protective cDNA clones are able to  
CC prevent, delay or reverse progression through the apoptotic or necrotic  
CC pathways when injected into a cell predisposed to or undergoing cell  
CC death. The cell death protective nucleic acids and polypeptides can be  
CC used in the diagnosis and treatment of disorders associated with cell  
CC death, and to screen for compounds which modulate their activity or  
CC expression. Such modulators, preferably a small organic molecule, an  
CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
CC cell death-related diseases. Such diseases include those associated with  
CC the central nervous system including psychiatric or neurological  
CC disorders, especially ischaemia-related conditions such as strokes, and  
CC also includes neurodegenerative disorders such as Alzheimer's disease,  
CC Huntington's disease, or Parkinson's disease. The modulators may also be  
CC used to treat infections such as meningitis, malaria, or trypanosomiasis;

cell death-associated disorder; central nervous system disorder; psychiatric disorder; neurological disorder; ischaemia-related disorder; stroke; cerebral infarction; ischaemic encephalopathy; neurodegenerative disorder; Alzheimer's disease; Huntington's disease; Parkinson's disease; infection; meningitis; malaria; trypanosomiasis; vascular disease; ophthalmological disorder; diabetic retinopathy; macular degeneration; hypertension; myocardial infarction; atherosclerosis; respiratory disorder; asthma; transgenic animal; chronic obstructive pulmonary disease; neoplastic condition; cancer; benign tumour; anaemia; gastrointestinal disorder; gastritis; ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder; glomerulonephritis; cystitis; endometriosis; endocrine disorder; Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis; urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS; open reading frame; ORF; ss.

Homo sapiens.

WO200145638-A2.

28-JUN-2001.

11-DEC-2000; 2000WO-US33547.

14-DEC-1999; 99US-0461697.

(COGE-) COGENT NEUROSCIENCE INC.

Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC; WPI; 2001-390297/41.

P-PSDB; AAG98654.

Novel protective sequence polynucleotides and polypeptides, used to identify modulators of their expression and activity, which are used in to treat central nervous system conditions, diseases and disorders -

Claim 2; Fig 6k; 325pp; English.

Sequences AAH84132-AAH84370 represent human nucleic acid sequences which protect against cell death (i.e., apoptosis or necrosis). Sequences AAH84132, AAH84145, AAH84170, AAH84201, AAH84226, AAH84285, AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones, while the remaining nucleic acid sequences within the range given above represent the open reading frames (ORFs) of these cDNA clones. Sequences AAG98610-AAG98829 represent the polypeptides encoded by the cell death protective ORFs. The cell death protective cDNA clones are able to prevent, delay or reverse progression through the apoptotic or necrotic pathways when injected into a cell predisposed to or undergoing cell death. The cell death protective nucleic acids and polypeptides can be used in the diagnosis and treatment of disorders associated with cell death, and to screen for compounds which modulate their activity or expression. Such modulators, preferably a small organic molecule, an antibody, a ribozyme, or an antisense molecule, can also be used to treat cell death-related diseases. Such diseases include those associated with the central nervous system including psychiatric or neurological disorders, especially ischaemia-related conditions such as strokes, and also includes neurodegenerative disorders such as Alzheimer's disease, Huntington's disease, or Parkinson's disease. The modulators may also be used to treat infections such as meningitis, malaria, or trypanosomiasis; vascular diseases such as ischaemic encephalopathy or cerebral infarction; eye conditions such as diabetic retinopathy or macular degeneration; hypertension; myocardial infarction; atherosclerosis; respiratory conditions such as asthma or chronic obstructive pulmonary disease; neoplastic conditions such as cancers or benign tumours; blood cell conditions such as anaemia; gastrointestinal conditions such as gastritis or ulcerative colitis; liver conditions such as biliary cirrhosis; kidney disorders such as glomerulonephritis; cystitis; endometriosis; endocrine disorders such as Grave's disease or Hashimoto's thyroiditis; skin conditions such as dermatitis or urticaria; or immune system disorders such as acquired immunodeficiency syndrome (AIDS). The nucleic acids may additionally be used to generate animal models of cell death-associated disorders. The present sequence represents a

cell death protective ORF.

Sequence 510 BP; 135 A; 116 C; 142 G; 117 T; 0 other;

Query Match 29.8%; Score 510; DB 22; Length 510; Best Local Similarity 100.0%; Pred. No. 2.2e-131; Indels 0; Matches 510; Conservative 0; Mismatches 0;

QY 1076 ATGCTACAGATTAGTAAGCAACTCTCTGAAGAATAATGAAGGATTGTCAATCCTCAGAG 1135

DB 1 ATGCTACAGATTAGTAAGCAACTCTCTGAAGAATAATGAAGGATTGTCAATCCTCAGAG 60

QY 1136 GCCACAGAGAGCGCTAAACCTCTGTGAAGATCAAGAGAACCTGTGAGGACATCACTTTT 1195

DB 61 GCCACAGAGAGCGCTAAACCTCTGTGAAGATCAAGAGAACCTGTGAGGACATCACTTTT 120

QY 1196 CCTGTCACTGAGGAGCTGGAGGCTGACCTTCTCTGAGAGACCACTCACTGCCTATGGGA 1255

DB 121 CCTGTCACTGAGGAGCTGGAGGCTGACCTTCTCTGAGAGACCACTCACTGCCTATGGGA 180

QY 1256 GTGCTTGGGGCTCAGAGCGAAGCGCTTCCATCTTAACCTGGAGTTGAAGCTTTCACACAG 1315

DB 181 GTGCTTGGGGCTCAGAGCGAAGCGCTTCCATCTTAACCTGGAGTTGAAGCTTTCACACAG 240

QY 1316 GCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGAATCTGCCCCATGTGAAATGGAG 1375

DB 241 GCTTCAAGTGCAGAGGTAATGCTTCTCTCTTTTGAATCTGCCCCATGTGAAATGGAG 300

QY 1376 CCTCAAGAAGTGAAGAAGGCAATCTCTCTGGGCATGTGCTGGGCAGTCACTGCTCTTC 1435

DB 301 CCTCAAGAAGTGAAGAAGGCAATCTCTCTGGGCATGTGCTGGGCAGTCACTGCTCTTC 360

QY 1436 GAGGAGCCTATGTCAAGCATGAGTGAAGTGGGATTCCTCAGAGCCCTGATGACTCAGAT 1495

DB 361 GAGGAGCCTATGTCAAGCATGAGTGAAGTGGGATTCCTCAGAGCCCTGATGACTCAGAT 420

QY 1496 AGCAGCTATGTTCCCACTCCACTCCACTCCACTCCACTCCACTCCACTCCACTCCACTCC 1555

DB 421 AGCAGCTATGTTCCCACTCCACTCCACTCCACTCCACTCCACTCCACTCCACTCCACTCC 480

QY 1556 CGCTCAAGAGAGGATGAGGAAATATAA 1585

DB 481 CGCTCAAGAGAGGATGAGGAAATATAA 510

RESULT 11

AAC08410

ID AAC08410 standard; cDNA; 544 BP.

AC AAC08410;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 12485.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST ) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1: SEQ ID 12485; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
SQ Sequence 544 BP; 122 A; 164 C; 137 G; 120 T; 1 other;

Query Match 27.4%; Score 468; DB 21; Length 544;  
Best Local Similarity 99.0%; Pred. No. 1.1e-119;  
Matches 483; Conservative 0; Mismatches 0; Indels 5; Gaps 1;  
QY 1 ACGGATCCTTGCTCAGGCTCTCGAGGTCGAGAGCCGCGCCGCGCTCTGCGAAG 60  
DB 57 ACAGGATCCTTGCTCAGGCTCTCGAGGTCGAGAGCCGCGCCGCGCTCTGCGAAG 116  
QY 61 CAGCAGTGAATAGTGTGCTACCTCTGCTCGGTCAGGTCAGACCTCCCGCTCTTCC 120  
DB 117 CAGCAGTGAATAGTGTGCTACCTCTGCTCGGTCAGGTCAGACCTCCCGCTCTTCC 176  
QY 121 GGCTGCCCTGAACGTCAGGAGCCTCAGGACCTGTGATTGGGCTCCGCTGTTCTTCG 180  
DB 177 GGCTGCCCTGAACGTCAGGAGCCTCAGGACCTGTGATTGGGCTCCGCTGTTCTTCG 236  
QY 181 GTGACCGAGAAACCCCTGAGGACCTGAGGACCTGAGGACCTGAGGACCTGAGGAC 240  
DB 237 GTGACCGAGAAACCCCTGAGGACCTGAGGACCTGAGGACCTGAGGACCTGAGGAC 296  
QY 241 TGCTCCTTTTCG-----GGCAAGGATCTCATTATCAGTCTTTGACCGACACAGAAATGCC 295  
DB 297 TGCTCCTTTTCGTTAAAGGCAAGATCTCATTATCAGTCTTTGACCGACACAGAAATGCC 356  
QY 296 TGGCATTGTATAATGTTTGTGAAGTGAAGACATATGGAAGTGAAGTGAAGTGAAGTGAAG 355  
DB 357 TGGCATTGTATAATGTTTGTGAAGTGAAGACATATGGAAGTGAAGTGAAGTGAAGTGAAG 416  
QY 356 TACTGGGAGAGATACCAATATCATCAAGCCAGCAGACAGAGTCTTCGATTGCTC 415  
DB 417 TACTGGGAGAGATACCAATATCATCAAGCCAGCAGACAGAGTCTTCGATTGCTC 476  
QY 416 CCACGGGAGTTCGCTGTGGTGAAGTCCATGACCCCTGACCAACCCCTCAGCCCAAC 475  
DB 477 CCACGGGAGTTCGCTGTGGTGAAGTCCATGACCCCTGACCAACCCCTCAGCCCAAC 536  
QY 476 AAGCCGAA 483  
DB 537 AAGCCGAA 544

RESULT 12  
AAK89991  
ID AAK89991 standard; DNA: 12244 BP.  
XX  
AC AAK89991;  
XX  
DT 05-NOV-2001 (first entry)  
XX

DE Human digestive system antigen genomic sequence SEQ ID NO: 3567.  
XX  
KW Human: digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.



stroke; cerebral infarction; ischaemic encephalopathy;  
neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
vascular disease; ophthalmological disorder; diabetic retinopathy;  
macular degeneration; hypertension; myocardial infarction;  
atherosclerosis; respiratory disorder; asthma; transgenic animal;  
chronic obstructive pulmonary disease; neoplastic condition; cancer;  
benign tumour; anaemia; gastrointestinal disorder; gastritis;  
ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS;  
open reading frame; ORF; ss.  
Homo sapiens.  
WO200145638-A2.  
28-JUN-2001.  
11-DEC-2000; 2000WO-US33547.  
14-DEC-1999; 99US-0461697.  
(COGE-) COGENT NEUROSCIENCE INC.  
Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
WPI: 2001-390297/41.  
P-PSDB; AAG98656.  
Novel protective sequence polynucleotides and polypeptides, used to  
identify modulators of their expression and activity, which are used in  
to treat central nervous system conditions, diseases and disorders -  
Claim 2; Fig 6M; 325pp; English.  
Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
protect against cell death (i.e., apoptosis or necrosis). Sequences  
AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
while the remaining nucleic acid sequences within the range given above  
represent the open reading frames (ORFs) of these cDNA clones. Sequences  
AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
protective ORFs. The cell death protective cDNA clones are able to  
prevent, delay or reverse progression through the apoptotic or necrotic  
pathways when injected into a cell predisposed to or undergoing cell  
death. The cell death protective nucleic acids and polypeptides can be  
used in the diagnosis and treatment of disorders associated with cell  
death, and to screen for compounds which modulate their activity or  
expression. Such modulators, preferably a small organic molecule, an  
antibody, a ribozyme, or an antisense molecule, can also be used to treat  
cell death-related diseases. Such diseases include those associated with  
the central nervous system including psychiatric or neurological  
disorders, especially ischaemia-related conditions such as strokes, and  
also includes neurodegenerative disorders such as Alzheimer's disease,  
Huntington's disease, or Parkinson's disease. The modulators may also be  
used to treat infections such as meningitis, malaria, or trypanosomiasis;  
vascular diseases such as ischaemic encephalopathy or cerebral  
infarction; eye conditions such as diabetic retinopathy or macular  
degeneration; hypertension; myocardial infarction; atherosclerosis;  
respiratory conditions such as asthma or chronic obstructive pulmonary  
disease; neoplastic conditions such as cancers or benign tumours; blood  
cell conditions such as anaemia; gastrointestinal conditions such as  
gastritis or ulcerative colitis; liver conditions such as biliary  
cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
system disorders such as acquired immunodeficiency syndrome (AIDS). The  
nucleic acids may additionally be used to generate animal models of  
cell death-associated disorders. The present sequence represents a  
cell death protective ORF.

SQ Sequence 336 BP; 84 A; 78 C; 96 G; 78 T; 0 other;  
Query Match 19.6%; Score 336; DB 22; Length 336;  
Best Local Similarity 100.0%; Pred. No. 4.2e-83;  
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1250 ATGGGAGTGTGGGGCTCAGAGCGAAGCGTTCCTCCATCTAACCTGGAGGTTGAAGCTTCA 1309  
DB 1 ATGGGAGTGTGGGGCTCAGAGCGAAGCGTTCCTCCATCTAACCTGGAGGTTGAAGCTTCA 60  
QY 1310 CCACAGGCTCAAGTCAGAGGTAATGCTTCTCTCTTTGGGAATCTGCCCATGTGAAA 1369  
DB 61 CCACAGGCTCAAGTCAGAGGTAATGCTTCTCTCTTTGGGAATCTGCCCATGTGAAA 120  
QY 1370 ATGGAGCCTCAAGAAAGTGAAGAGCAATGTCTCTGGGCATGTGTCTGGGCAGTGAT 1429  
DB 121 ATGGAGCCTCAAGAAAGTGAAGAGCAATGTCTCTGGGCATGTGTCTGGGCAGTGAT 180  
QY 1430 GTCTTCGAGGAGCCTATGTGAGGCATGTGAGTGAAGCTTCTCAGAGCCCTGATGAC 1489  
DB 181 GTCTTCGAGGAGCCTATGTGAGGCATGTGAGTGAAGCTTCTCAGAGCCCTGATGAC 240  
QY 1490 TCAGATAGCAGCTATGGTCCCACTCCACTGACAGCCTCATGGGGTCTTCCCTGTTTC 1549  
DB 241 TCAGATAGCAGCTATGGTCCCACTCCACTGACAGCCTCATGGGGTCTTCCCTGTTTC 300  
QY 1550 AACCCAGCGTGCAGAGAGGAGGATGAGGAAAAATATAA 1585  
DB 301 AACCCAGCGTGCAGAGAGGAGGATGAGGAAAAATATAA 336  
RESULT 14  
ABK41954  
ID ABK41954 standard; CDNA; 517 BP.  
XX AC ABK41954;  
XX 21-MAY-2002 (first entry)  
DE CDNA encoding novel human connective tissue related polypeptide #342.  
XX KW Human; connective tissue related disorder; cancer; gene therapy;  
XX KW cytotstatic; gene; ss.  
OS Homo sapiens.  
XX PN WO200155343-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01322.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-565190/63.  
DR P-PSDB; AAU86776.  
XX  
XX Nucleic acid encoding novel connective tissue associated polypeptides,  
PT used in diagnosing, preventing, treating or ameliorating a disorder  
PT such as cancer or rheumatoid arthritis -  
XX  
PS Claim 4; SEQ ID No 352; 673pp; English.  
XX  
XX The present invention relates to the isolation of novel human connective  
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide  
CC (cDNA and genomic) sequences encoding them. The sequences of the  
CC invention are useful in the diagnosis, treatment, prevention and/or  
CC prognosis of diseases associated with connective tissue(s), including  
CC cancer. The polynucleotide sequences of the invention are also useful  
CC in gene therapy. ABK41613-ABK42101 represent cDNA sequences encoding  
CC the novel human connective tissue related polypeptides.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 517 BP; 108 A; 140 C; 132 G; 133 T; 4 other;  
Query Match 17.2%; Score 293.8; DB 23; Length 517;  
Best Local Similarity 96.1%; Pred. No. 2.8e-71;  
Matches 321; Conservative 1; Mismatches 9; Indels 3; Gaps 2;



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:57:18 ; Search time 59.8436 Seconds  
(without alignments)  
821.584 Million cell updates/sec

Title: US-09-857-308-1

Perfect score: 2175

Sequence: 1 MNLQRYWGEIPISSQTNRS.....SLMGSPVFNQRCCKMRKI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/FCI\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep1.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep2.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep3.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2155	99.1	412	11	US-09-922-261-77
2	1898	87.3	363	11	US-09-922-261-81
3	1022	47.0	199	11	US-09-922-261-93
4	858	39.4	169	11	US-09-922-261-97
5	572	26.3	111	11	US-09-922-261-101
6	371	17.1	71	11	US-09-922-261-105
7	239	11.0	46	11	US-09-922-261-113
8	224	10.3	43	11	US-09-922-261-115
9	118	5.4	377	10	US-09-879-957-36
10	115	5.3	783	10	US-09-888-615-67
11	112	5.1	486	10	US-09-922-217-685
12	112	5.1	486	11	US-09-833-263-685
13	112	5.1	486	15	US-10-025-380-685
14	110.5	5.1	785	11	US-09-801-368-348
15	106	4.9	478	15	US-10-156-634A-2
16	106	4.9	599	11	US-09-758-269-6

17	106	4.9	1182	15	US-10-156-634A-4	Sequence 4, Appli
18	103.5	4.8	2224	15	US-10-115-563-14	Sequence 14, Appl
19	103.5	4.8	2224	15	US-10-172-712-31	Sequence 31, Appl
20	102.5	4.7	930	15	US-10-113-794A-1	Sequence 1, Appli
21	102.5	4.7	1243	15	US-10-196-935A-4	Sequence 4, Appli
22	101	4.6	504	15	US-10-029-180-6	Sequence 6, Appli
23	99.5	4.6	568	15	US-10-136-960-8	Sequence 8, Appli
24	99.5	4.6	663	12	US-09-932-257A-21	Sequence 21, Appli
25	99.5	4.6	730	12	US-09-932-257A-23	Sequence 23, Appli
26	99	4.6	656	10	US-09-825-301-9	Sequence 9, Appli
27	99	4.6	656	10	US-09-759-143-379	Sequence 379, App
28	99	4.6	656	10	US-09-780-669-379	Sequence 379, App
29	99	4.6	656	10	US-09-810-936-305	Sequence 305, App
30	99	4.6	656	10	US-09-822-827-379	Sequence 379, App
31	99	4.6	656	11	US-09-429-755-305	Sequence 305, App
32	99	4.6	656	11	US-09-924-400-305	Sequence 305, App
33	99	4.6	656	11	US-09-895-793-379	Sequence 379, App
34	99	4.6	656	11	US-09-895-814-379	Sequence 379, App
35	99	4.6	656	15	US-10-012-896-379	Sequence 379, App
36	99	4.6	656	15	US-10-010-940-379	Sequence 379, App
37	99	4.6	656	15	US-10-212-679-305	Sequence 305, App
38	99	4.6	864	11	US-09-883-096-2	Sequence 2, Appli
39	99	4.6	1719	10	US-09-759-143-378	Sequence 378, App
40	99	4.6	1719	10	US-09-780-669-378	Sequence 378, App
41	99	4.6	1719	10	US-09-822-827-378	Sequence 378, App
42	99	4.6	1719	11	US-09-895-793-378	Sequence 378, App
43	99	4.6	1719	11	US-09-895-814-378	Sequence 378, App
44	99	4.6	1719	15	US-10-012-896-378	Sequence 378, App
45	99	4.6	1719	15	US-10-010-940-378	Sequence 378, App

ALIGNMENTS

RESULT 1

US-09-922-261-77  
; Sequence 77, Application US/09922261  
; Patent No. US20020111471A1  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/922,261  
; CURRENT FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US/09/461,697  
; PRIOR FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 77  
; LENGTH: 412  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-261-77

Query Match	99.18;	Score	2155;	DB 11;	Length	412;			
Best Local Similarity	100.0%;	Pred. No.	6.9e-185;						
Matches	410;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	5	RWGEIPISSQTNRSFDLLPREFRLVEVHDPLHQPSANKPKPPTMLDIPSECSLTI	64						

```
QY 125 AFFQIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 184
Db 123 AFFQIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 182
QY 185 VAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSLQKFWOHRKIDYHS 244
Db 183 VAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSLQKFWOHRKIDYHS 242
QY 245 YMLQISKOLSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSELEADLASGDOSLPM 304
Db 243 YMLQISKOLSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSELEADLASGDOSLPM 302
QY 305 GVLGAQSERFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHGVLSGDV 364
Db 303 GVLGAQSERFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSEEGNVSGHGVLSGDV 362
QY 365 FEPMGSMSEAGIPQSPDDSDSYGSHSTDSLGMSSPVFNQCKRMKI 414
Db 363 FEPMGSMSEAGIPQSPDDSDSYGSHSTDSLGMSSPVFNQCKRMKI 412

RESULT 2
US-09-922-261-81
; Sequence 81, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-81

Query Match 87.3%; Score 1898; DB 11; Length 363;
Best Local Similarity 100.0%; Pred. No. 6.5e-162;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 MLDIPSECSLTITHTLIQHNRRLNLIATAQAOQOQTEGVKTESEPLPSCGSPPL 111
Db 1 MLDIPSECSLTITHTLIQHNRRLNLIATAQAOQOQTEGVKTESEPLPSCGSPPL 60
QY 112 PDDLPLDCKNPAPFOIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGF 171
Db 61 PDDLPLDCKNPAPFOIRHSDPESDFYRGKEPVTLSWHSCROLLYQAVATILAHAGF 120
QY 172 DCANESVLETLTDVAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSL 231
Db 121 DCANESVLETLTDVAHEYCLKFTKLLRAVDREARLGOTPPDVMEQVFHEVGIGSVLSL 180
QY 232 QKFWOHRKIDYHSMQISKOLSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSEEL 291
Db 181 QKFWOHRKIDYHSMQISKOLSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSEEL 240
QY 292 EADLASGDOSLPMGVLGAQSERFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSE 351
Db 241 EADLASGDOSLPMGVLGAQSERFFSNLEVEASPOASSAEVNASPLWNLAHVKMEPQSE 300
```

```
QY 352 GNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSYGSHSTDSLGMSSPVFNQCKRMKI 411
Db 301 GNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSYGSHSTDSLGMSSPVFNQCKRMKI 360
QY 412 RKI 414
Db 361 RKI 363

RESULT 3
US-09-922-261-93
; Sequence 93, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-93

Query Match 47.0%; Score 1022; DB 11; Length 199;
Best Local Similarity 100.0%; Pred. No. 1e-83;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 MEQVFHEVGIGSVLSLQKFWOHRKIDYHSMQISKOLSEYERIVNPEKATEDAKPVKI 275
Db 1 MEQVFHEVGIGSVLSLQKFWOHRKIDYHSMQISKOLSEYERIVNPEKATEDAKPVKI 60
QY 276 KEEPVS DITFPVSELEADLASGDOSLPMGVLGAQSERFFSNLEVEASPOASSAEVNASP 335
Db 61 KEEPVS DITFPVSELEADLASGDOSLPMGVLGAQSERFFSNLEVEASPOASSAEVNASP 120
QY 336 LWNLAHVKMEPQSEEGNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSYGSHSTDS 395
Db 121 LWNLAHVKMEPQSEEGNVSGHGVLSGVDFEPMGSMSEAGIPQSPDDSDSYGSHSTDS 180
QY 396 LMGSSPVFNQCKRMKI 414
Db 181 LMGSSPVFNQCKRMKI 199

RESULT 4
US-09-922-261-97
; Sequence 97, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
```

```

; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-97

Query Match          39.4%; Score 858; DB 11; Length 169;
Best Local Similarity 100.0%; Pred. No. 4.1e-69;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 MLQISKQLSEYERIVNPKATEDAKPVKIKEEPVSDITFPVSEELADLASGDSLPMG 305
DB 1 MLQISKQLSEYERIVNPKATEDAKPVKIKEEPVSDITFPVSEELADLASGDSLPMG 60

QY 306 VLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSDVF 365
DB 61 VLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSDVF 120

QY 366 EEPMSGMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKRMKI 414
DB 121 EEPMSGMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKRMKI 169

RESULT 5
US-09-922-261-101
; Sequence 101, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-101

Query Match          26.3%; Score 572; DB 11; Length 111;
Best Local Similarity 100.0%; Pred. No. 1e-43;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 MGVLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSD 363
DB 1 MGVLGAQSERFPNLEVEASPOASSAEVNASPLNLAHVKNMEPQSEEGNVSGHVLGSD 60

QY 364 VFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKRMKI 414
DB 61 VFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKRMKI 111

RESULT 6
US-09-922-261-105
; Sequence 105, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-105

Query Match          17.1%; Score 371; DB 11; Length 71;
Best Local Similarity 100.0%; Pred. No. 5.5e-26;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 344 MEQSEEGNVSGHVLGSDVFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVF 403
DB 1 MEQSEEGNVSGHVLGSDVFEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVF 60

QY 404 NQCKRMKI 414
DB 61 NQCKRMKI 71

RESULT 7
US-09-922-261-113
; Sequence 113, Application US/09922261
; Patent No. US20020111471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-113

Query Match          11.0%; Score 239; DB 11; Length 46;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 MSGMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNQCKRMKI 414
```

```
Db 1 MSGMSAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQRCCKRMKI 46
|||||
RESULT 8
US-09-922-261-115
; Sequence 115, Application US/09922261
; Patent No. US2002011471A1
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/922,261
; PRIOR FILING DATE: 2001-08-03
; PRIOR FILING DATE: US/09/461,697
; PRIOR FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-261-115

Query Match 10.3%; Score 224; DB 11; Length 43;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 MSEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQRCCKRMKI 414
|||||
Db 1 MSEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQRCCKRMKI 43
|||||

RESULT 9
US-09-879-957-36
; Sequence 36, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOWLKES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
```

Query Match 5.3%; Score 115; DB 10; Length 783;  
Best Local Similarity 21.1%; Pred. No. 0.19;  
Matches 109; Conservative 61; Mismatches 178; Indels 168; Gaps 25;

QY 32 VEVDHPLP-SANKPKPTMLDIPSE-----PCSLTIHTIQLIQRRLRL 79  
DB 1 MRVQDPTKALPEKAKSKRPT---VPHDESSDIAVGLTCQHVSHAISV-----NH 49  
QY 80 IATAOONOOOTRGVTESEPLPCPGSPPLPDDL-LPLDC-----KNPNAPFOIRHS 132  
DB 50 VKRAIAEN---LWSVSECELEERFYDQGLVLTSDIWLCLKFCQCGKNSQSOLKH- 105  
QY 133 DPESDYRKGEK-----VTELSWHSCROLLYQAVATILAH-----169  
DB 106 -----FKSSRTEPHCIILNSTWIWCYECDEKLSTHCNKKVLAQIVDFLQKHASKTQTS 160  
QY 170 -----GFCANESV-----LETLTDVAHE 188  
DB 161 AFSRIMKLCEKEKETEIOKGGKCRNLRSVRGINTGLNCTCFNNAVMQNLAQTYTLTDLMNE 220  
QY 189 YCLKFTKLLRF-----AVDREARLGOTPPDVMEQVFFH---EVGIGSVLSLQKFWQ 236  
DB 221 IKESSTKLTFFPSDSQLOPLVLSRPG--PLTSALFLFLHSMKTEKGPLSPKVLFNQ 278  
QY 237 -----HRIKDYHMYLQISKQL-----SEYERIV-----NPEKATED---AK 271  
DB 279 LCQKAPRFKDFQO---QDSQELLHYLLDAVRTTEETRIQASILKAFNNPTTKTADDETRK 335  
QY 272 PVKIK--EEPVSIDITEPVSEE--LEADLASQDSQSLPMGVLAGQASERFSPNLEVE--ASPQA 326  
DB 336 KVAISTVKOPFIDISLPIIEERVSERVKPLLMGRNMKNYSLRETDHRYSGNVTIENIHOPRA 395  
QY 327 SSAEVNASPLNLAHVK-----MEPQSEEGNVSCHGVLSGSDVFEEMPMSGMSEA 375  
DB 396 AKHSSSKDKSLIHLDBKCIKRLKSGSETVYQKNENLNGDSLMFASLMN-----SES 449  
QY 376 GIPQSPDDSSSYSGSHSTDSL-MGSPVFNQRCR 410  
DB 450 RLNESPTDDSEKASHESNVNADSEPSSESASKQ 485

RESULT 11  
US-09-922-217-685  
; Sequence 685, Application US/09922217  
; Patent No. US20020076414A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Smith, Carole Lynn  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 21021.471C13  
; CURRENT APPLICATION NUMBER: US/09/922,217  
; CURRENT FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 685  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-217-685

Query Match 5.1%; Score 112; DB 10; Length 486;  
Best Local Similarity 22.4%; Pred. No. 0.17;

Matches 56; Conservative 33; Mismatches 83; Indels 78; Gaps 9;

QY 192 KFTKLLRFAVDREARLGOTPPDVMEQVFEHVGIGSVLSLQKFWQHRKIDYHMYLQISK 251  
DB 207 KYEKS-----KELDOGTPOYMEMEQVFEQ-----CQOFEKRLRFFREVLLEVK 253  
QY 252 QLS-----EYERIVNPEKATEDAKPVKIKEEPVSDITFPVSELEADL-----295  
DB 254 HLNLSNVAGYKAIYHDLQESIRAADAVEDLRFWRANHGPGMAMNWPQFEWSADLIRTLS 313  
QY 296 -----ASDQSLPMGVLAGQASERFSPNLEVEASPOASAEVNASPLWN 338  
DB 314 RREKKATDGTFTLTGINOTGDQFLP-----SKPSTLNVPSNP-AQSAQSS-----360  
QY 339 LAHVKMEPQSEEGNVSCHGVLSGSDVFEEP-----MSGMSEAGIPQSPDDSDSSYVSH 391  
DB 361 -----YNPEDEDDT-----GSTVSEKEDIKAKNVSSYEKTSQSYPTDWSDDSDSNPPS 408  
QY 392 STDLSMGSSP 401  
DB 409 STDANGDSNP 418

RESULT 12  
US-09-833-263-685  
; Sequence 685, Application US/09833263  
; Patent No. US20020110547A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 21021.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 685  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-833-263-685

Query Match 5.1%; Score 112; DB 11; Length 486;  
Best Local Similarity 22.4%; Pred. No. 0.17;  
Matches 56; Conservative 33; Mismatches 83; Indels 78; Gaps 9;

QY 192 KFTKLLRFAVDREARLGOTPPDVMEQVFEHVGIGSVLSLQKFWQHRKIDYHMYLQISK 251  
DB 207 KYEKS-----KELDOGTPOYMEMEQVFEQ-----CQOFEKRLRFFREVLLEVK 253  
QY 252 QLS-----EYERIVNPEKATEDAKPVKIKEEPVSDITFPVSELEADL-----295  
DB 254 HLNLSNVAGYKAIYHDLQESIRAADAVEDLRFWRANHGPGMAMNWPQFEWSADLIRTLS 313  
QY 296 -----ASDQSLPMGVLAGQASERFSPNLEVEASPOASAEVNASPLWN 338  
DB 314 RREKKATDGTFTLTGINOTGDQFLP-----SKPSTLNVPSNP-AQSAQSS-----360  
QY 339 LAHVKMEPQSEEGNVSCHGVLSGSDVFEEP-----MSGMSEAGIPQSPDDSDSSYVSH 391  
DB 361 -----YNPEDEDDT-----GSTVSEKEDIKAKNVSSYEKTSQSYPTDWSDDSDSNPPS 408  
QY 392 STDLSMGSSP 401  
DB 409 STDANGDSNP 418

RESULT 13  
US-10-025-380-685  
; Sequence 685, Application US/10025380

Publication No. US20020182191A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Lodes, Michael J.  
APPLICANT: Secrist, Heather  
APPLICANT: Benson, Darin R.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Stolk, John A.  
APPLICANT: Wang, Tongtong  
APPLICANT: Jiang, Yugu  
APPLICANT: Smith, Carole L.  
APPLICANT: King, Gordon E.  
APPLICANT: Wang, Aljun  
APPLICANT: Clapper, Jonathan D.  
APPLICANT: Skeiky, Yasir A. W.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedwick, Thomas S.  
APPLICANT: Carter, Darrick  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
FILE REFERENCE: 210121.471C14  
CURRENT APPLICATION NUMBER: US/10/025,380  
CURRENT FILING DATE: 2001-12-19  
NUMBER OF SEQ ID NOS: 1129  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 685  
LENGTH: 486  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-025-380-685

Query Match 5.1%; Score 112; DB 15; Length 486;  
Best Local Similarity 22.4%; Pred. No. 0.17;  
Matches 56; Conservative 33; Mismatches 83; Indels 78; Gaps 9;  
QY 192 KFTLLRFAVREARLGOTPPDWEQVHEVGVLSLQKFWQHRKDYHSHYMLQISK 251  
DB 207 KYEKL-----KELDQGTPOYMNMEQVFEQ-----COQFEKRLRFREVLLEVQK 253  
QY 252 OLS-----EYERIVNPEKATEDAKPVKIKEPVS DITFPVSEELADL----- 295  
DB 254 HNLNSVAGYKAIYHDLQSTRADAVEDLRFRANHGPGMANNWPQFEWSADLIRTLS 313  
QY 296 -----ASGDQLPMGVLGASERFPNSNLEVEASPOASSAEVNASPLWN 338  
DB 314 RREKKATDGTLTGINTGTGQFLP-----SKPSSTLNVSPN-AQSAQSQSS----- 360  
QY 339 LAHVMEQSEBEGNVSGHVLGSDVFREP-----MSGMSEAGIQSPDPDSDSSYSGH 391  
DB 361 -----YNPFEDEDDT-----GSTVSEKEDIKARNVSSYEKTSQYPTDWSDESNNPFS 408  
QY 392 STDLSMGSSP 401  
DB 409 STDANGSNP 418

RESULT 14  
US-09-801-368-348  
Sequence 348, Application US/09801368  
Patent No. US20020128250A1  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Cali, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff

APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/801,368  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: US 09/487,558  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: US 60/160,587  
PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 348  
LENGTH: 785  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-348

Query Match 5.1%; Score 110.5; DB 11; Length 785;  
Best Local Similarity 18.8%; Pred. No. 0.48;  
Matches 106; Conservative 67; Mismatches 189; Indels 201; Gaps 24;  
QY 8 GEIPSSSOTNRSSFDL--LPREFRLVEVHDPPLHQ-----PSANKPKPPTMLDI 55  
DB 247 GSMSPNSNRTSRNSNSISSLAQPPMSNYQPPSTYQYPGFKHTSSIPNSHSPIPRSLT 306  
QY 56 PSE-----PCSLTIHTIQLIQRRLRLNLIAQAQNOQOOTEVGVKTESEPLPSCPGS 108  
DB 307 PTGPTSQNGPLSYNLQVGLL-----PPOQQQVSPLYDGNISITPPVKPST 353  
QY 109 PP-----LPDDLPLDCKNPNAPFO---LRHSDPESDFYRGKE----- 144  
DB 354 DQETYLTAHRHGVSDDQYDSMAKTMS--FQTTIRHPMPLIATTNATGNSNTSGTSASIR 412  
QY 145 -PVTLSWHSCROLLYQAVA---FILAHAGFDCANESVLETLTDVAHEYCLKFTKLLRFA 200  
DB 413 PRVTTTWEDEKTLCYQVEANGISVVRADNDWNGTKL-----LNVTKMTRGR 461  
QY 201 VDREARLGOTPPDVMQVHEVGVLSLQKFW-----QHRIDKY----- 242  
DB 462 RDGILK-----AEKIRHVVKIGS-MHLKGWIPFERALATAQREKIADYLYPLFIR 511  
QY 243 --HSMYLOIS-----KOLSEE--YERIVN----- 262  
DB 512 DIQSVLKQNNPNSDSSSSSTGIKISIPRTYQYFINNYQNPNGPNSIAAQLTYSSMNL 571  
QY 263 -----PEKATEDAKPVKIKEPVS DITFPVSEELADLASGQSL-----PM--GVLG--A 309  
DB 572 NKKIIPNNSIPAVTIAAGEKPLKCKTNPNSNQLEGHITITLQTLSATMPMKQOLMGNA 631  
QY 310 QSERFPNSNLEVEASPOASSAEVNASPLWNLAHVMEQSEBEGNV----- 354  
DB 632 SPLSYPRNATMNSASTLGITPADSKPL-TPSPTTTNTNQSSSNVGSVHTGTLPRVESE 690  
QY 355 -----SCHGVLGSDVFEEPMSGM-----SEAGIQSPDPD 383  
DB 691 SASHSKWSKEADSGNTYPDNQTLEKPRSSQLPISALTSTDTKIKTSTDEATQPNPSE 750  
QY 384 SDSSYSGSHSTDLSM-GSSPVFNQ 405  
DB 751 AEPVKESESQSDYDAGDYSNE 773

RESULT 15  
US-10-156-634A-2  
Sequence 2, Application US/10156634A  
Publication No. US20030013152A1  
GENERAL INFORMATION:  
APPLICANT: DeJong, Jeff L.  
TITLE OF INVENTION: Transcription Factors Related to TFIIA  
FILE REFERENCE: 119941-1092  
CURRENT APPLICATION NUMBER: US/10/156,634A  
CURRENT FILING DATE: 2002-05-28  
PRIOR APPLICATION NUMBER: 09/326,529



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:55:28 ; Search time 31.3934 Seconds  
(without alignments)  
388.015 Million cell updates/sec

Title: US-09-857-308-1

Perfect score: 2175

Sequence: 1 MNLQRYWGEIPISSTQNRK.....SLMGSSPVFNQRCKRMKI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2155	99.1	412	4	US-09-461-697-77
2	1898	87.3	363	4	US-09-461-697-81
3	1022	47.0	199	4	US-09-461-697-93
4	858	39.4	169	4	US-09-461-697-97
5	572	26.3	111	4	US-09-461-697-101
6	371	17.1	71	4	US-09-461-697-105
7	239	11.0	46	4	US-09-461-697-113
8	224	10.3	43	4	US-09-461-697-115
9	123	5.7	552	1	US-07-999-280A-22
10	123	5.7	552	1	US-08-426-279-22
11	123	5.7	552	1	US-08-401-013-22
12	123	5.7	552	3	US-08-426-570-22
13	123	5.7	552	3	US-08-425-876-22
14	123	5.7	552	4	US-08-426-243-22
15	123	5.7	552	4	US-08-401-632-22
16	118	5.4	377	4	US-08-630-915A-36
17	114	5.2	552	1	US-07-999-280A-24
18	114	5.2	552	1	US-08-426-279-24
19	114	5.2	552	1	US-08-401-013-24
20	114	5.2	552	3	US-08-426-570-24
21	114	5.2	552	3	US-08-425-876-24
22	114	5.2	552	4	US-08-426-243-24
23	114	5.2	552	4	US-08-401-632-24
24	106	4.9	478	4	US-09-326-529-2
25	106	4.9	1182	4	US-09-326-529-4
26	105	4.8	501	2	US-08-969-630-4
27	99	4.6	656	4	US-09-605-785-379

28	99	4.6	656	4	US-09-439-313-379	Sequence 379, App
29	99	4.6	656	4	US-09-352-616A-379	Sequence 379, App
30	99	4.6	1719	4	US-09-605-785-378	Sequence 378, App
31	99	4.6	1719	4	US-09-439-313-378	Sequence 378, App
32	99	4.6	1719	4	US-09-352-616A-378	Sequence 378, App
33	98.5	4.5	1342	1	US-07-978-895-4	Sequence 4, Appli
34	98.5	4.5	1342	2	US-08-473-119-4	Sequence 4, Appli
35	98.5	4.5	1342	2	US-08-475-352-4	Sequence 4, Appli
36	98.5	4.5	1343	6	5183884-4	Patent No. 5183884
37	98.5	4.5	1618	1	US-07-853-913-4	Sequence 4, Appli
38	98	4.5	822	4	US-08-684-932A-38	Sequence 38, Appl
39	97.5	4.5	1234	2	US-08-317-10A-15	Sequence 15, Appl
40	97.5	4.5	1234	5	PCT-US95-13041-15	Sequence 15, Appl
41	97.5	4.5	1243	2	US-08-557-139-2	Sequence 2, Appli
42	97.5	4.5	1342	2	US-08-484-438-9	Sequence 9, Appli
43	96.5	4.4	502	2	US-08-969-630-2	Sequence 2, Appli
44	96	4.4	1464	1	US-08-026-138E-1	Sequence 1, Appli
45	96	4.4	1706	2	US-08-459-568-2	Sequence 2, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-461-697-77  
Sequence 77, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENT NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Puranam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461.697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 77  
LENGTH: 412  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-461-697-77

Query Match 99.1%; Score 2155; DB 4; Length 412;  
Best Local Similarity 100.0%; Pred. No. 4.9e+205;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	5	RYNGEIPISSTQNRSSFDLLPREFLVEVHDPLHPHOSANKPKPTMLDIPSEPCSLTI	64
Db	3	RYNGEIPISSTQNRSSFDLLPREFLVEVHDPLHPHOSANKPKPTMLDIPSEPCSLTI	62
Qy	65	HTIQLIHNRLRLNLATAQAQQOQTEGVKTESEPLPSCGSPPLPDDLLPLDCKNPN	124
Db	63	HTIQLIHNRLRLNLATAQAQQOQTEGVKTESEPLPSCGSPPLPDDLLPLDCKNPN	122
Qy	125	APQIRHSDPESDFYRGKGPEVTVLSHWSHSCROLLYQAVATILAHAGDCANESVLEFLTD	184
Db	123	APQIRHSDPESDFYRGKGPEVTVLSHWSHSCROLLYQAVATILAHAGDCANESVLEFLTD	182
Qy	185	VAHEYCLKFTLLRFAVDFREARLGOTFPDVMQVFEVGVIGSVLSLQKFWQHRKDYHS	244
Db	183	VAHEYCLKFTLLRFAVDFREARLGOTFPDVMQVFEVGVIGSVLSLQKFWQHRKDYHS	242
Qy	245	YMLQISKQLSEEVERIVNPEKATEDAKPVKIKEEPVSDITFPVSEELADLASGDSLPM	304
Db	243	YMLQISKQLSEEVERIVNPEKATEDAKPVKIKEEPVSDITFPVSEELADLASGDSLPM	302

QY 305 GVLGAQSERFSPNLEVEASQASSAEVNASPLWNLNLAHVKMEPQSEEGNVSCHGVLSGV 364  
Db 303 GVLGAQSERFSPNLEVEASQASSAEVNASPLWNLNLAHVKMEPQSEEGNVSCHGVLSGV 362  
QY 365 FEEPMSGSEAGIPQSPDDSSYSGSHSTDSLMGSSPVFNQCKRMKI 414  
Db 363 FEEPMSGSEAGIPQSPDDSSYSGSHSTDSLMGSSPVFNQCKRMKI 412

## RESULT 2

US-09-461-697-81  
; Sequence 81, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 81  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-461-697-81

Query Match 87.3%; Score 1898; DB 4; Length 363;  
Best Local Similarity 100.0%; Pred. No. 1.1e-179;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 52 MLDIPSEPCSLTIHTIQLIQRNRLNLIATAQANQOQTEGVKTESEPLPSCPGSPPL 111  
Db 1 MLDIPSEPCSLTIHTIQLIQRNRLNLIATAQANQOQTEGVKTESEPLPSCPGSPPL 60  
QY 112 PDDLPLDCKNPAPFQIRHSDPESDFYRGKGEVPTLSWHSQRLLYQAVATILAHAGF 171  
Db 61 PDDLPLDCKNPAPFQIRHSDPESDFYRGKGEVPTLSWHSQRLLYQAVATILAHAGF 120  
QY 172 DCANESVLETLTDVAHEYCLFETKLLRFAVDREARLGOTPPDVMEQVFEVHGVLSL 231  
Db 121 DCANESVLETLTDVAHEYCLFETKLLRFAVDREARLGOTPPDVMEQVFEVHGVLSL 180  
QY 232 QKFWQHRKDYHSMQLISKOLSEYERIVNPERKATEDAKPVKIKEEPPVSDITPPVSEEL 291  
Db 181 QKFWQHRKDYHSMQLISKOLSEYERIVNPERKATEDAKPVKIKEEPPVSDITPPVSEEL 240  
QY 292 EADLASGQDQSLPMGVGAQSERFSPNLEVEASQASSAEVNASPLWNLNLAHVKMEPQSEEE 351  
Db 241 EADLASGQDQSLPMGVGAQSERFSPNLEVEASQASSAEVNASPLWNLNLAHVKMEPQSEEE 300  
QY 352 GNVSGHVLGSDVDFEPPMSGSEAGIPQSPDDSSYSGSHSTDSLMGSSPVFNQCKRM 411  
Db 301 GNVSGHVLGSDVDFEPPMSGSEAGIPQSPDDSSYSGSHSTDSLMGSSPVFNQCKRM 360  
QY 412 RKI 414  
Db 361 RKI 363

## RESULT 3

US-09-461-697-93  
; Sequence 93, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:

; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 93  
; LENGTH: 199  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-461-697-93

Query Match 47.0%; Score 1022; DB 4; Length 199;  
Best Local Similarity 100.0%; Pred. No. 2.6e-93;  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 216 MEQVFHEVIGSVLSLQKFWQHRKDYHSMQLISKOLSEYERIVNPERKATEDAKPVKI 275  
Db 1 MEQVFHEVIGSVLSLQKFWQHRKDYHSMQLISKOLSEYERIVNPERKATEDAKPVKI 60  
QY 276 KEEPVSITFPVSEELADLASGQDQSLPMGVGAQSERFSPNLEVEASQASSAEVNASP 335  
Db 61 KEEPVSITFPVSEELADLASGQDQSLPMGVGAQSERFSPNLEVEASQASSAEVNASP 120  
QY 336 LWNLAHVKMEPQSEEGNVSCHGVLSGVDFEPPMSGSEAGIPQSPDDSSYSGSHSTDS 395  
Db 121 LWNLAHVKMEPQSEEGNVSCHGVLSGVDFEPPMSGSEAGIPQSPDDSSYSGSHSTDS 180  
QY 396 LMGSSPVFNQCKRMKI 414  
Db 181 LMGSSPVFNQCKRMKI 199

## RESULT 4

US-09-461-697-97  
; Sequence 97, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: Lo, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Puranam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 97  
; LENGTH: 169  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-461-697-97

Query Match 39.4%; Score 858; DB 4; Length 169;  
Best Local Similarity 100.0%; Pred. No. 3.5e-77;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 246 MLQISKOLSEYERIVNPERKATEDAKPVKIKEEPPVSDITFPVSEELADLASGQDQSLPMG 305

```
Db 1 MLIQISQLSSEYRIINPEKATDAPVKIKEEPVSDITFPVSEELADIASGDQSLPMG 60
QY 306 VLGAQSERFNSLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSDVF 365
Db 61 VLGAQSERFNSLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSDVF 120
QY 366 EEPMSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVFNQCKKMRKI 414
Db 121 EEPMSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVFNQCKKMRKI 169

RESULT 5
US-09-461-697-101
; Sequence 101, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-101

Query Match 26.3%; Score 572; DB 4; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.8e-49;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 MGVGAQSERFNSLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSD 363
Db 1 MGVGAQSERFNSLEVEASPOASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSD 60
QY 364 VFEPMMSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVFNQCKKMRKI 414
Db 61 VFEPMMSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVFNQCKKMRKI 111

RESULT 6
US-09-461-697-105
; Sequence 105, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105
; LENGTH: 71
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-105

Query Match 17.1%; Score 371; DB 4; Length 71;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 344 MEQSEEGNVSGHVLGSDVFEPMMSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVF 403
Db 1 MEQSEEGNVSGHVLGSDVFEPMMSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVF 60

QY 404 NQCKKMRKI 414
Db 61 NQCKKMRKI 71

RESULT 7
US-09-461-697-113
; Sequence 113, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-113

Query Match 11.0%; Score 239; DB 4; Length 46;
Best Local Similarity 100.0%; Pred. No. 9e-17;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 MSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVFNQCKKMRKI 414
Db 1 MSGMSEAGIQSPDDSDSSYSGSHSTDLSLMGSSPVFNQCKKMRKI 46

RESULT 8
US-09-461-697-115
; Sequence 115, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115
```

```
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-115

Query Match      10.3%; Score 224; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 372 MSEAGIQSPDSDSSYSGSHSTDSLMGSSPVFNQCKRMRKI 414
      |||||
Db 1 MSEAGIQSPDSDSSYSGSHSTDSLMGSSPVFNQCKRMRKI 43

RESULT 9
US-07-999-280A-22
; Sequence 22, Application US/07999280A
; Patent No. 5573930
; GENERAL INFORMATION:
; APPLICANT: LADNER, MARTHA B.
; APPLICANT: NOBLE, JANELLE A.
; APPLICANT: MARTIN, GEORGE A.
; APPLICANT: KAWASAKI, ERNEST S.
; APPLICANT: COYNE, MAZIE YEE
; APPLICANT: HALENBECK, ROBERT F.
; APPLICANT: KOTHS, KIRSTON E.
; TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/999,280A
; FILING DATE: 28-DEC-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGarrigle Jr., Philip L.
; REGISTRATION NUMBER: 31,395
; REFERENCE/DOCKET NUMBER: 0681.007
; TELEPHONE: (510) 601-2718
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-999-280A-22

Query Match      5.7%; Score 123; DB 1; Length 552;
Best Local Similarity 21.6%; Pred. No. 0.0016;
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

Qy 14 SSQTNRSFDLLPREFLRVEVHDPLHQPSANKPKPTMLDIPSEPCSLTHTLIQHN 73
      |||||
Db 226 SORTEGSS--LLPSELPL-----RIEDPGSAKORP-----PRSTCQ-TLESTEQPNHG 270
      |||||
Qy 74 RRLNLIATAQAQNOQTEGVKTESEPLPSCPCSPPLPDLLPLDCKNPAPQIRHSD 133
      |||||
Db 271 DRL-----TEDSQHPGAGGVPVGVEDIL-----294

Qy 134 PESDFYRGKGFVTELSWHSCRQLLYQAVATILAHAGFDCAVESVLETITDVAHEYCLKF 193
```

```
Db 295 -ESSL--GTNNVLEEASGESEGLTQ-----EAKFSFSTPVGGS-----331
      |||||
Qy 194 TKLIRFAVDREARLGQTPFP-----DYNQOVFHEVG-----IGSVLSLQKFWQHRK 240
      |||||
Db 332 ---IQAETDRPRALSASFPPKSTEDQKPVDTIDRPLTEVNPMPRIGQT-----QNTTP 381
      |||||
Qy 241 DYHSYMLQISKQJSEYERIVNPEKATEDAKPVKIKEEPVSDITFPVSEEELEADLASDQ 300
      |||||
Db 382 E-----KTDGTSRLREDHQEGSPHIATPNPQVNSNSATPVAQLLLPKSH-----SWG 430
      |||||
Qy 301 SLPNGVL---GAQSERFPNSLVEASPOSS---AEVNASPLNLAHVAKMEPOESEGN 353
      |||||
Db 431 VLPGLGELEGKRSTRDRRSPAELEGSSASEGAARPVARNFNSIPLTDTGHV-----EQHEG- 484
      |||||
Qy 354 VSGHGVLSGSDVFEPPMSG-----MSEAGI-----PQSPDDSDSSYSGSHSTDSL 396
      |||||
Db 485 -SSDPQIPESVFHLLVFGIILVLTGGLLFYKWKWRSHRDPQTLDDSSVGRPESSSL 540
      |||||

RESULT 10
US-08-426-279-22
; Sequence 22, Application US/08426279
; Patent No. 5672343
; GENERAL INFORMATION:
; APPLICANT: LADNER, MARTHA B.
; APPLICANT: NOBLE, JANELLE A.
; APPLICANT: MARTIN, GEORGE A.
; APPLICANT: KAWASAKI, ERNEST S.
; APPLICANT: COYNE, MAZIE YEE
; APPLICANT: HALENBECK, ROBERT F.
; APPLICANT: KOTHS, KIRSTON E.
; TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,279
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: McGarrigle Jr., Philip L.
; REGISTRATION NUMBER: 31,395
; REFERENCE/DOCKET NUMBER: 0681.013
; TELEPHONE: (510) 601-2718
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-426-279-22

Query Match      5.7%; Score 123; DB 1; Length 552;
Best Local Similarity 21.6%; Pred. No. 0.0016;
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

Qy 14 SSQTNRSFDLLPREFLRVEVHDPLHQPSANKPKPTMLDIPSEPCSLTHTLIQHN 73
      |||||
Db 226 SORTEGSS--LLPSELPL-----RIEDPGSAKORP-----PRSTCQ-TLESTEQPNHG 270
```



```

; INFORMATION FOR SEQ ID NO: 22:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 552 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-426-570-22

```

Query Match 5.7%; Score 123; DB 3; Length 552;  
Best Local Similarity 21.6%; Pred. No. 0.0016;  
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

RESULT 13  
US-08-425-876-22  
; Sequence 22, Application US/08425876  
; Patent No. 6117422  
; GENERAL INFORMATION:  
; APPLICANT: LADNER, MARTHA B.  
; APPLICANT: NOBLE, JANELLE A.  
; APPLICANT: MARTIN, GEORGE A.  
; APPLICANT: KAWASAKI, ERNEST S.  
; APPLICANT: COYNE, MAZIE YEE  
; APPLICANT: HALENBECK, ROBERT F.  
; APPLICANT: KOTHS, KIRSTON E.  
; TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: Intellectual Property - R440, P.O. Box 8037  
; CITY: Emeryville  
; STATE: California

```

; REGISTRATION NUMBER: 31,395
; REFERENCE/DOCKET NUMBER: 0681.011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2718
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PS-08-425-876-22

```

Query Match.	5.7%;	Score 123;	DB 3;	Length 552;
Best Local Similarity	21.6%;	Pred. No. 0.0016;		
Matches	90;	Conservative	47;	Mismatches 144; Indels 136; Gaps 20;
QY	14	SSOTNRSSFLLLDPRFRFLVVDHPPHLHPSANKPKPTMLDIPSECLSIHTIQLIQHN	73	
DB	226	SQTEGSS--LLSELP--	226	RIEDPSAKQRP-----PRSTCQ-TLESTEQPNHG
QY	74	RLRLNLATAQAQNOQQTGKVTSEEPLEPSCFSPPLDPDILLPLDCKNPNAPFQIRHSD	133	
DB	271	DRL-----TEDSQHPHSAGGVPGVEDIL-----	294	
QY	134	PESDFYRGKCEPVTELSWHSCROLLQVATILAHAGFDCANESVLETLTDVAHEYCLKF	193	
DB	295	ESSL--GTNNVLEEASGESEGLTQ-----EAKTSPSPVGVGS-----	331	
QY	194	TKLLRFADVREARLGOTPPF-----DVMEQVFHEVG-----IGSVLSIQKFWQHRIK	240	
DB	332	---IQAETDRPRALSASPPKSTEDQKPDVITDRPLTEVNPMPIGOT-----QNTPT	381	
QY	241	DYHSYMLQISKQJSEBYERVINPEKATDAKPKVKEEPPVSDTTFPVSELEADLASGDQ	300	
DB	382	E-----KTDGTSTLRDHQEGSPHIATPNPQRVSNSATPVQAQLLPKSH-----SWG	430	
QY	301	SLPMGVL---GAQSERFPNLEVASPOASS---AEVNASPLWNLAHVKMEPOESSEGN	353	
DB	431	VLPFGELEGKRTDRRSPAELEGGEAGAPVAFNSIPTDTGCHV-----EQHEG-	484	
QY	354	VSGHGVLGSDVFEEPMSG-----MSEAGI-----PQSPDDSSSYGSHSTDLSL	396	
DB	485	SSDPOTPRSEVFHLLVPGIILVLLTVGAILFYKKWKWESHDPOTLDSSVGRPEBSSSL	540	

RESULT 14  
US-08-426-243-22  
; Sequence 22, Application US/08426243  
; Patent No. 6146851  
; GENERAL INFORMATION:  
; APPLICANT: LADNER, MARTHA B.  
; APPLICANT: NOBLE, JANELLE A.  
; APPLICANT: MARTIN, GEORGE A.  
; APPLICANT: KAWASAKI, ERNEST S.  
; APPLICANT: COYNE, MAZIE YEE  
; APPLICANT: HALENBECK, ROBERT F.  
; APPLICANT: KOTHS, KIRSTON E.  
; TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: Intellectual Property - R440, P.O. Box 8097  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0. Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,243  
FILING DATE: 28-DEC-1992  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 07/999,280  
FILING DATE: 28-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGarrigle Jr., Philip L.  
REGISTRATION NUMBER: 31,395  
REFERENCE/DOCKET NUMBER: 0681.007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2718  
TELEFAX: (510) 655-3542  
TELEX: n/a  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 552 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-426-243-22

Query Match 5.7%; Score 123; DB 4; Length 552;  
Best Local Similarity 21.6%; Pred. No. 0.0016;  
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

QY 14 SSOTNRSSFDLLPREFRLVEHDPPLHQPSSANKPKPTMLDIPSEPCSLTIHTIQLIQHN 73  
DB 226 SORTEGSS--LLPSELPL-----RIEDPGSAKQRP-----PRSTCQ-TLESTEQPNHG 270  
QY 74 RRLNRLIATAQAQNOQQTGKTESEPLPSCGSPPLDPLDCKNPNAPFQIRHSD 133  
DB 271 DRL-----TEDSQPHSAGGPVGVEDIL----- 294  
QY 134 PSDFYRGKGPVTELSWHSRQLLYQAVATILAHAGFCANESVLETLTDVAHEYCLKF 193  
DB 295 -ESSL--GTNWLEASGEASEGFLTQ-----EAKFSPSTPVGGS----- 331  
QY 194 TKLLFAVDREARLGOTPP-----DVMEQVFEVG-----IGSVLSLQKFWOHRK 240  
DB 332 ---IQAETDRPRALSAPPKSTEDQKPVDTIDRPLTEVNPMPICQT-----QNNTP 381  
QY 241 DYHSYMLQISKOLSEYERIVNPEKATEDAKPVKIKEEVPVSDITFPVSEELADLASGQ 300  
DB 382 E----KTDGTSTLRDHQEPGSPHIATPNPQVNSATPVAQLLPKSH-----SWG 430  
QY 301 SLPMGVL-----GAQSERFFSNLEVEASPOASS---AEVNASPLWNLAHVMEPOESEEGN 353  
DB 431 VLPLGELEGRSTRDRRSAPAELEGGASGEAARFVARNFNSIPLTDGTV-----EQHEG- 484  
QY 354 VSGHGVLSGVDFEPMGSG-----MSEAGI-----PQSPDSDSSYSGSHSTDLSL 396  
DB 485 -SSDQIPESVFHLLVPGIILVLLTVGGLLFYKWKWRSHRDPQTLDSVGRPEDSSL 540

RESULT 15  
US-08-401-632-22  
Sequence 22, Application US/08401632  
Patent No. 6204020  
GENERAL INFORMATION:  
APPLICANT: LADNER, MARTHA B.  
APPLICANT: NOBLE, JANELIE A.  
APPLICANT: MARTIN, GEORGE A.  
APPLICANT: KAWASAKI, ERNEST S.  
APPLICANT: COYNE, MAZIE YEE  
APPLICANT: HALENBECK, ROBERT F.  
APPLICANT: KOTHS, KIRSTON E.  
TITLE OF INVENTION: NEW FORMS OF COLONY STIMULATING FACTOR-1  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: Intellectual Property - R440, P.O. Box 8097

CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/401,632  
FILING DATE: 09-MAR-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: McGarrigle Jr., Philip L.  
REGISTRATION NUMBER: 31,395  
REFERENCE/DOCKET NUMBER: 0681.009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2718  
TELEFAX: (510) 655-3542  
TELEX: n/a  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 552 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-401-632-22

Query Match 5.7%; Score 123; DB 4; Length 552;  
Best Local Similarity 21.6%; Pred. No. 0.0016;  
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

QY 14 SSOTNRSSFDLLPREFRLVEHDPPLHQPSSANKPKPTMLDIPSEPCSLTIHTIQLIQHN 73  
DB 226 SORTEGSS--LLPSELPL-----RIEDPGSAKQRP-----PRSTCQ-TLESTEQPNHG 270  
QY 74 RRLNRLIATAQAQNOQQTGKTESEPLPSCGSPPLDPLDCKNPNAPFQIRHSD 133  
DB 271 DRL-----TEDSQPHSAGGPVGVEDIL----- 294  
QY 134 PSDFYRGKGPVTELSWHSRQLLYQAVATILAHAGFCANESVLETLTDVAHEYCLKF 193  
DB 295 -ESSL--GTNWLEASGEASEGFLTQ-----EAKFSPSTPVGGS----- 331  
QY 194 TKLLFAVDREARLGOTPP-----DVMEQVFEVG-----IGSVLSLQKFWOHRK 240  
DB 332 ---IQAETDRPRALSAPPKSTEDQKPVDTIDRPLTEVNPMPICQT-----QNNTP 381  
QY 241 DYHSYMLQISKOLSEYERIVNPEKATEDAKPVKIKEEVPVSDITFPVSEELADLASGQ 300  
DB 382 E----KTDGTSTLRDHQEPGSPHIATPNPQVNSATPVAQLLPKSH-----SWG 430  
QY 301 SLPMGVL-----GAQSERFFSNLEVEASPOASS---AEVNASPLWNLAHVMEPOESEEGN 353  
DB 431 VLPLGELEGRSTRDRRSAPAELEGGASGEAARFVARNFNSIPLTDGTV-----EQHEG- 484  
QY 354 VSGHGVLSGVDFEPMGSG-----MSEAGI-----PQSPDSDSSYSGSHSTDLSL 396  
DB 485 -SSDQIPESVFHLLVPGIILVLLTVGGLLFYKWKWRSHRDPQTLDSVGRPEDSSL 540

Search completed: July 18, 2003, 10:01:22  
Job time : 34.3934 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:03:57 ; Search time 25.5071 Seconds  
(without alignments)  
673.193 Million cell updates/sec

Title: US-09-857-308-1

Perfect score: 2175

Sequence: 1 MNLQRYWGEIPISSTQNR.....SLMGSSPVFNQCKRMKRI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	123	5.7	552	1 CSF1_MOUSE	P07141 mus musculus
2	117	5.4	486	1 PAC2_HUMAN	Q9unf0 homo sapien
3	117	5.4	575	1 SP6D_BACSU	P37963 bacillus su
4	112.5	5.2	1332	1 SP77_YEAST	P35177 saccharomyc
5	110.5	5.1	785	1 SOK2_YEAST	P33438 saccharomyc
6	109	5.0	486	1 PAC2_MOUSE	Q9wve8 mus musculus
7	108.5	5.0	671	1 CHS5_YEAST	Q12114 saccharomyc
8	106.5	4.9	1859	1 GBF1_HUMAN	Q92538 homo sapien
9	106	4.9	387	1 INVO_TARBA	P24711 tarsius ban
10	106	4.9	478	1 T2AY_HUMAN	Q9unn4 homo sapien
11	106	4.9	1182	1 T2A2_HUMAN	Q9y6g2 homo sapien
12	105.5	4.9	4377	1 ANK3_HUMAN	Q12955 homo sapien
13	105	4.8	488	1 PAC2_RAT	Q9qy17 rattus norv
14	105	4.8	896	1 Y249_HUMAN	Q92539 homo sapien
15	103.5	4.8	1324	1 SAL1_HUMAN	Q9nsc2 homo sapien
16	103.5	4.8	2224	1 FA5_HUMAN	P12259 homo sapien
17	103	4.7	830	1 VPP3_HUMAN	Q13488 h vacuolar
18	103	4.7	1462	1 NKCR_HUMAN	P30414 homo sapien
19	102.5	4.7	3358	1 PGCV_MOUSE	Q62059 mus musculus
20	102	4.7	1461	1 TOP2_CANAL	P87078 candida alb
21	102	4.7	3381	1 PGCV_BOVIN	P81282 bos taurus
22	101	4.6	317	1 APE_PAPAN	P05770 papio anubi
23	100.5	4.6	434	1 VATB_BORBU	O51120 borrelia bu
24	100.5	4.6	901	1 Y298_HUMAN	O15016 homo sapien
25	100	4.6	1153	1 PVDH_PLANK	P50493 plasmodium
26	99.5	4.6	466	1 IMP2_DROME	P29681 drosophila
27	99.5	4.6	476	1 HLX1_MOUSE	Q61670 mus musculus
28	99.5	4.6	568	1 DCPY_ZYMO	P06672 zymomonas m
29	99.5	4.6	730	1 CCT2_HUMAN	O60583 homo sapien
30	99.5	4.6	1095	1 NEB1_RAT	O35867 rattus norv
31	99.5	4.6	1185	1 MAPX_DROME	P23226 drosophila
32	99.5	4.6	1235	1 IRS1_RAT	P35570 rattus norv
33	98.5	4.5	682	1 AMPH_CHICK	P50478 gallus gall

#### RESULT 1

ID	CSF1_MOUSE	STANDARD;	PRT;	552 AA.
AC	P07141;			
DT	01-APR-1988 (Rel. 07, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Macrophage colony stimulating factor-1 precursor (CSF-1) (MCSF).			
GN	CSF1 OR CSFM.			
OS	Mus musculus (Mouse), and			P21860 homo sapien
OS	Rattus norvegicus (Rat).			Q49429 mycoplasma
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			P48681 homo sapien
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			Q9nzm4 homo sapien
OX	NCBI_TaxID=10090, 10116;			Q9y6d6 homo sapien
RN	[1]			Q15911 homo sapien
RP	SEQUENCE FROM N.A.			Q26417 strongyloce
RP	SPECIES=Mouse;			Q14774 homo sapien
RX	MEDLINE=87174763; PubMed=3494232;			P35569 mus musculus
RA	Delamarter J.F., Hession C., Semon D., Gough N.M., Rothenbuhler R.,			P35568 homo sapien
RA	Mermoud J.-J.;			Q13796 homo sapien
RT	"Nucleotide sequence of a cDNA encoding murine CSF-1			
RT	(Macrophage-CSF).";			
RL	Nucleic Acids Res. 15:2389-2390(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	SPECIES=Mouse;			
RX	MEDLINE=88320507; PubMed=2457916;			
RA	Ladner M.B., Martin G.A., Noble J.A., Wittman V.P., Warren M.K.,			
RA	McGrogan M., Stanley E.R.;			
RT	"cDNA cloning and expression of murine macrophage colony-stimulating			
RT	factor from L929 cells.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 85:6706-6710(1988).			
RN	[3]			
RP	SEQUENCE OF 1-100 FROM N.A.			
RP	SPECIES=Mouse;			
RX	MEDLINE=85242709; PubMed=3493488;			
RA	Rajavashisth T.B., Eng R., Shaddock R.K., Waheed A., Ben-Avram C.M.,			
RA	Shively J.E., Lusis A.J.;			
RT	"Cloning and tissue-specific expression of mouse macrophage colony-			
RT	stimulating factor mRNA.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 84:1157-1161(1987).			
RN	[4]			
RP	SEQUENCE OF 33-57.			
RP	SPECIES=Mouse;			
RX	MEDLINE=85242709; PubMed=3925458;			
RA	Ben-Avram C.M., Shively J.E., Shaddock R.K., Waheed A.,			
RA	Rajavashisth T.B., Lusis A.J.;			
RT	"Amino-terminal amino acid sequence of murine colony-stimulating			
RT	factor 1.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:4486-4489(1985).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RP	SPECIES=rat; STRAIN=Wistar; TISSUE=Muscle;			
RX	MEDLINE=93363632; PubMed=8357831;			
RA	Borycki A.G., Lenormund J., Guillier M., Leibovitch S.A.;			
RT	"Isolation and characterization of a cDNA clone encoding for rat			

#### ALIGNMENTS

CSF-1 gene. Post-transcriptional repression occurs in myogenic differentiation.";  
RL Biochim. Biophys. Acta 1174:143-152(1993).  
RN [6]  
RP SEQUENCE OF 1-13 FROM N.A.  
RC SPECIES-Mouse;  
RX MEDLINE-91340149; PubMed-1874443;  
RA Harrington M.A., Edenberg H.J., Saxman S.M., Pedigo L.M., Daub R.,  
RA Broxmeyer H.E.;  
RT Cloning and characterization of the murine promoter for the colony-stimulating factor-1-encoding gene.";  
RL Gene 102:165-170(1991).  
CC -!- FUNCTION: GRANULOCYTE/MACROPHAGE COLONY-STIMULATING FACTORS ARE CYTOKINES THAT ACT IN HEMATOPOIESIS BY CONTROLLING THE PRODUCTION, DIFFERENTIATION, AND FUNCTION OF 2 RELATED WHITE CELL POPULATIONS OF THE BLOOD. THE GRANULOCYTES AND THE MONOCYTES-MACROPHAGES.  
CC -!- FUNCTION: THIS CSF INDUCES MACROPHAGES  
CC -!- SUBUNIT: HOMODIMER, LINKED BY THREE DISULFIDE BONDS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X05010; CAA28660.1; -  
DR EMBL; M21952; AAA37481.1; -  
DR EMBL; M21149; AAA37482.1; -  
DR EMBL; M15692; AAA37480.1; -  
DR EMBL; M84361; AAA03032.1; -  
DR EMBL; M81316; AAA19866.1; -  
DR PIR; A23166; A23166  
DR PIR; A26575; A26575  
DR PIR; A31401; A31401.  
DR MGB; MGI:1339753; Csf1.  
KW Cytokine; Growth factor; Glycoprotein; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 552 MACROPHAGE COLONY STIMULATING FACTOR-1.  
FT DISULFID 63 63 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 39 122 BY SIMILARITY.  
FT DISULFID 80 171 BY SIMILARITY.  
FT DISULFID 134 178 BY SIMILARITY.  
FT DISULFID 189 189 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 191 191 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 292 292 D -> G.  
FT VARIANT 345 345 S -> P.  
FT CONFLICT 3 3 MISSING (IN REF. 3).  
FT CONFLICT 6 6 A -> R (IN REF. 3).  
FT CONFLICT 7 8 AG -> PR (IN REF. 3).  
FT CONFLICT 246 246 P -> A (IN REF. 1).  
SQ SEQUENCE 552 AA; 60648 MW; 3886D72D70E770AF CRC64;  
  
Query Match  
Best Local Similarity 5.7%; Score 123; DB 1; Length 552;  
Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;  
  
QY 14 SSQTNRSFDLLPREFLVEVHDPLHQPANSKPKPTMLDIPSEPCSLTIHTLIQHN 73  
DB 226 SQTEGSS--LLPSLPL-----RIEDPGSAKQRP-----PRSTCO-TLESTEQPNHG 270  
QY 74 RRLRLNLTAAQAOQOQTEGVKTESEPLPCSGSPPLDPLDCKNPANPFOIRUSD 133  
DB 271 DRL-----TEDSOPHPSAGGVPVGVEDIL----- 294  
QY 134 PESDFYRGKGPVTELSWHSCROLLYQAVAILAHAGFDCAVESLFTLTDVAHEYCLKF 193  
DB 295 -ESSL--GTNVLEASGEASEGLTQ-----EAKFSPTPVGGS----- 331

QY 194 TKLLRFVAVDREARLQOTFPF-----DYMEOVFHEVG-----IGSYLSLQKFWQHRK 240  
DB 332 --IQAETDRPRALSASPEPKSTEDQKPVDTIDRPLTEVNMRRPQT-----QNTP 381  
QY 241 DYHSYMLQISKOLSEERYIVNPEKATEDAKPVKIKEEPVSDITFPVSEELADLASGDQ 300  
DB 382 E-----KTDGTSTLRDHQEGSPHIATPNPQVNSATPFAQLLLPKSH-----SWG 430  
QY 301 SLPMGVL----GAOSERFPPSLEVEASPOASS---AEVNASPLNLAHVMEPOESERGN 353  
DB 431 VLPGLGELEKGRSTRDRRSPALEGGSASEGARVPARENSIPLTDTGHV-----EQHEG- 484  
QY 354 VSGHVLGSDVFEPMVG-----MSEAGI-----PQSPDDSDSYGSHSDSL 396  
DB 485 -SSDPQIPESVFLHLLVPGIILVLTGVLGFLFYKWKRSRHPDQTLDSVGRPESSSL 540  
  
RESULT 2  
PAC2\_HUMAN  
ID PAC2\_HUMAN STANDARD; PRT; 486 AA.  
AC Q9UNF0; Q9H0D3; Q96HV9; Q9NPNI; Q95921; Q9Y4V2;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Protein kinase C and casein kinase substrate in neurons protein 2.  
GN PACSIN2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
SE [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE-Retina, and Brain;  
RX MEDLINE-99358785; PubMed=10431838;  
RA Ritter B., Modregger J., Paulsson M., Plomann M.;  
RT "PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter proteins.";  
RL FEBS Lett. 454:356-362(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE-Testis;  
RX MEDLINE-21154917; PubMed=11230166;  
RA Wienann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S., Ansgore W., Boecher M., Bloecher H., Bauersachs S., Blum H., Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N., Mexes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D., Wambutt R., Korn B., Klein M., Poustka A.;  
RT "Towards a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs.";  
RL Genome Res. 11:422-435(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M., Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K., Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J., Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R., Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G., Evans K.L., Fey J.M., Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C., Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A., Laird G.K., Langford C.F., Levenson M.A., Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T., Mcclay J., McLaren S., McMurray A., Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T., RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,

RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chissee S., Murray J., Miller N., Minx P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley J., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton D., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,  
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim O.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,  
 RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tilahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=SKIN;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE OF 335-486 FROM N.A. (ISOFORM 2).  
 RA Auffray C., Ansong W., Ballabio A., Estivill X., Gibson K.,  
 RA Lehrach H., Poustka A., Lundeberg J.;  
 RT "The European IMAGE consortium for integrated molecular analysis of  
 human gene transcripts.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RP FUNCTION.  
 RX PubMed=11082044;  
 RA Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;  
 RT "All three PACSIN isoforms bind to endocytic proteins and inhibit  
 endocytosis.";  
 RL J. Cell Sci. 113:4511-4521(2000).  
 CC -!- FUNCTION: May play a role in vesicle formation and transport.  
 CC -!- SUBUNIT: Homo- and hetero-aggregates with other PACSINS. Binds  
 CC dynamin 1, synaptotagmin, synapsin 1 and the neural Wiskott-Aldrich  
 CC syndrome protein (N-WASP) (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic  
 CC distribution (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed.  
 CC -!- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C  
 CC (PKC) (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE PACSIN FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 FCH DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF128536; AAD41781.1; -;  
 DR EMBL; AL136845; CAB66779.1; -;  
 DR EMBL; AL022476; CAB39175.1; -;  
 DR EMBL; AL049758; CAB51395.1; -;  
 DR EMBL; BC008037; AAH08037.1; -;  
 DR EMBL; AL389984; CAB97538.1; -;  
 DR Genbank; HGNC:8571; PACSIN2.  
 DR MIM; 604960; -;

DR HSP; Q60631; IGBQ.  
 DR InterPro; IPR001060; Cdc15\_Fes\_CIP4.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00611; FCH; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR PRODOM; PD000066; SH3; 1.  
 DR SMART; SM00055; FCH; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS01033; FCH; 1.  
 DR PROSITE; PS00002; SH3; 1.  
 KW Endocytosis; SH3 domain; Coiled coil; Phosphorylation;  
 KW Alternative splicing.  
 FT DOMAIN 11 75 FCH.  
 FT DOMAIN 426 486 SH3.  
 FT DOMAIN 184 219 COILED COIL (POTENTIAL).  
 FT VARSPIC 344 384 MISSING (IN ISOFORM 2).  
 FT VARIANT 294 294 M -> I (IN DESNP:2746984).  
 FT /FTID=VAR\_013711.  
 FT VARIANT 324 324 V -> F (IN DBSNP:1062913).  
 FT /FTID=VAR\_013712.  
 FT CONFLICT 182 182 L -> F (IN REF. 1).  
 FT CONFLICT 256 256 D -> N (IN REF. 1).  
 FT CONFLICT 309 309 N -> I (IN REF. 1).  
 FT CONFLICT 336 336 S -> F (IN REF. 1).  
 FT CONFLICT 378 380 DDT -> EDI (IN REF. 1).  
 SQ SEQUENCE 486 AA; 55738 MW; 821DBEF65DAD1AA8 CRC64;  
 Query Match 5.4%; Score 117; DB 1; Length 486;  
 Best Local Similarity 22.8%; Pred. No. 0.67;  
 Matches 57; Conservative 32; Mismatches 83; Indels 78; Gaps 9;  
 QY 192 KFTKLLFAVDREARLQGPFDVMEQVFHEVIGSVLSLQKFWOHRKIDYHVMLOISK 251  
 Db 207 KYEKS-----KELDQGTPOYMNMEQVFEQ-----CQFEKKLRFREVLLEVK 253  
 QY 252 QLS-----EYERIVNPEKATEDAKPKKEPVSDITFPVSEELADL----- 295  
 Db 254 HLDLSNVAGYKAIYHDLQESIRADAVEDLRFRANHGPGAMNWPQFEWSADLNRTLS 313  
 QY 296 -----ASQDQLPMGVLAQSERFSPNLEVASFOASAEVNASPLWN 338  
 Db 314 RREKKKATDGVTLTGINTQDQLP-----SKPSTLNVPSPN-AQSAQSS----- 360  
 QY 339 LAHVMEPOPSSEGNVSGHGVLDVFEFP-----MSCMSEAGIPQSPDDSDSSVSGSH 391  
 Db 361 -----YNPFEEDDT-----GSTVSEKDDTKAKNVSSYEKTSQPTDMSDDSENPF 408  
 QY 392 STDLSMGSSP 401  
 Db 409 STDANGSNP 418  
 RESULT 3  
 SP6D\_BACSU STANDARD; PRT; 575 AA.  
 ID SP6D\_BACSU  
 AC P37963;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Stage VI sporulation protein D.  
 GN SPOVID.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP MEDLINE=93194796; PubMed=8449878;  
 RA Beall B., Driks A., Losick R., Moran C.P. Jr.;  
 RT "Cloning and characterization of a gene required for assembly of the  
 Bacillus subtilis spore coat.";  
 RL J. Bacteriol. 175:1705-1716(1993).  
 RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN=166;
RX MEDLINE=98044033; PubMed=9384377;
RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Brullis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchli C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chai S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferraci E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrin P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpatra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zunshtein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- FUNCTION: REQUIRED FOR ASSEMBLY OF A NORMAL SPORE COAT. MAY BE A
CC COMPONENT OF THE INNERMOST LAYER OF THE SPORE COAT THAT AIDS IN
CC ITS ADHERENCE TO THE PRESPORE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS INITIATED AROUND THE SECOND
CC HOUR OF SPOREALATION AND CONTINUES THROUGHOUT DEVELOPMENT. MAY BE
CC EXPRESSED PREDOMINANTLY IN THE MOTHER CELL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07792; AAA22808.1; -;
DR EMBL; Z99118; CAB14771.1; -;
DR Subtilist; BG10346; spovID.
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00257; LysM; 1.
KW Sporulation; Complete proteome.
SQ SEQUENCE 575 AA; 64976 MW; 9A879AB16B18884F CRC64;
Query Match 5.4%; Score 117; DB 1; Length 575;
Best Local Similarity 21.4%; Pred. No. 0.84; Indels 108; Gaps 20;
Matches 87; Conservative 53; Mismatches 158;
QY 9 EIPISQNTNRSSDLLPRFLVEVHDPPLH-----QPSANKKPPTML-DIPSEPCSLT 63
DQ 146 EIPLYEA-----PAAPREELSEPPAHSVVEEPGASSAEAVLQHEPPAEPPEL- 194
QY 64 IHTQLQHNRLRLNLTAAQNOQQTGKVTETSEPLSCPSPLPDDL----- 115
DQ 195 -----FISKAGLELETE-----KAESEPPESVASEPAREDKVEERSEE 236
QY 116 LPLCKNPAPFQIRHSDPED-----FYRCK---GEPTVLSWHSCROLLYQAVAT 164
DQ 237 LAVPETEVRASETESEPEPDPPSEIETQIVAKKETAPEAAIA--DVREEDADSPAET 294

```

```

165 IL-AHAG-----FDCANESVL-----EFLTDAVHEYCLKFTKLRFVDRRLGQT 210
295 ELRHVGAEESPAELHSETVIKKEETVSPNHEYALR-----QEAQNEEAAQSDQ 349
211 PFDV-MEVPHEVGISV-----LSLQK-----FWOHRKDYHVMQ-----IS 250
350 ADPALCOEAEAPDEALESVSAALSIEDSRETASAVYMENDNADLHFHFNKTSSEASQ 409
251 KQLSEEVERIVNPKATED---AKPVKIKPEVSDITFPVSELEADLASDQSLPQVGL 307
410 EELPEPAYRTLPQEEEDSFYSAPKLLLEEEQEESEFIEVRKTPSAEPEKKEPTFP--- 466
308 GAQSERPSPNLEVEASQASAEVNASPLMNLAHVKMPEQSESGN 353
467 --QSFLPESSETERK-----ETDAVPRVAPAAETKEPQTKESDN 504

```

RESULT 4

ID	SPT7_YEAST	STANDARD	PRT	1332 AA
AC	P35177			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Transcriptional activator SPT7.			
GN	SPT7 OR YBR081C OR YBR0739.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RC	MEDLINE=95229044; PubMed=7713415;			
RA	Gansheroff L.J., Dollard C., Tan P., Winston F.;			
RT	"The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein			
RT	important for transcription in vivo.";			
RL	Genetics 139:523-536(1995).			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RC	MEDLINE=95076715; PubMed=7985423;			
RA	van der Aart Q.J.M., Barthe C., Dolignon F., Aigle M., Crouzet M.,			
RA	Steenma H.Y.;			
RT	"Sequence analysis of a 31 kb DNA fragment from the right arm of			
RT	Saccharomyces cerevisiae chromosome II.";			
RL	Yeast 10:959-964(1994).			
RP	SEQUENCE OF 1-835 FROM N.A.			
RC	STRAIN=S288C;			
RC	MEDLINE=92285152; PubMed=1350857;			
RA	Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,			
RA	David I.B.;			
RT	"The bromodomain: a conserved sequence found in human, Drosophila and			
RT	yeast proteins.";			
RL	Nucleic Acids Res. 20:2603-2603(1992).			
CC	-!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY			
CC	OTHER GENES.			
CC	-!- SUBCELLULAR LOCATION: Nuclear.			
CC	-!- SIMILARITY: CONTAINS 1 BROMODOMAIN.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)



```

RESULT 6
PAC2_MOUSE
ID PAC2_MOUSE STANDARD; PRT; 486 AA.
AC Q9WVE8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons protein 2.
GN PACSIN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 x DBA;
RX MEDLINE=99358785; PubMed=10431838;
RA Ritter B., Modregger J., Paulsson M., Plomann M.;
RT "PACSIN 2, a novel member of the PACSIN family of cytoplasmic adapter
RT proteins.";
RL FEBS Lett. 454:356-362(1999).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
FUNCTION.
RX PubMed=11082044;
RA Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;
RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
RT endocytosis.";
RL J. Cell Sci. 113:4511-4521(2000).
CC -!- FUNCTION: May play a role in vesicle formation and transport.
CC -!- SUBUNIT: Homo- and hetero-aggregates with other PACSINs. Binds
CC dynamin 1, synaptotagmin, synapsin 1 and the neural Wiskott-Aldrich
CC syndrome protein (N-WASP) (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
CC distribution (By similarity).
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed, with highest levels in
CC brain, heart, skeletal muscle and ovaries.
CC -!- PM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
CC (PKC) (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 FCH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF128535; AAD41780.1; -
CC DR EMBL; BC023502; AAB23502.1; -
CC DR HSSP; P29355; 1SEM.
CC DR MGD; MGI:1345153; Pacsin2.
CC DR InterPro; IPR001060; Cdc15_Fes_CIP4.
CC DR InterPro; IPR001452; SH3.
CC DR Pfam; PF00611; FCH; 1.
CC DR Pfam; PF00018; SH3; 1.
CC DR PRINTS; PD000066; SH3; 1.
CC DR PRODOM; PD000066; SH3; 1.
CC DR SMART; SM00055; FCH; 1.
CC DR SMART; SM00326; SH3; 1.
CC DR PROSITE; PS00133; FCH; 1.
CC DR PROSITE; PS00002; SH3; 1.
CC DR Endocytosis; SH3 domain; Coiled coil; Phosphorylation.
CC FT DOMAIN 11 75 FCH.
CC FT DOMAIN 426 486 SH3.
CC FT DOMAIN 184 239 COILED COIL (POTENTIAL).

```

```

SQ SEQUENCE 486 AA; 55833 MW; 66C17ECC6767E0E7 CRC64;
Query Match 5.0%; Score 109; DB 1; Length 486;
Best Local Similarity 21.4%; Pred. No. 2.4;
Matches 69; Conservative 35; Mismatches 94; Indels 124; Gaps 13;
Oy 167 AHAGFDGANESVLETLTDV-----AHEYCLKFTKLLRFAVDREA----- 205
Db 134 AEDGFRKAQKPAWKKLKEVEAAKHAHTACKEEKL---AISREANSRADPSLNPEQLKKL 190
Oy 206 -----RLGQT--PFDPVMEOVFHEVGIGSVLSLQKFWQHRIK 240
Db 191 QDKTEKCKQDVLTKDKYKSLKELDTTPQYMEHMEQVFEQ-----CQFEKKLR 242
Oy 241 DYHSMQLQISKOLS-----EEYRTVNPKEKATAKPVKIKEEPPVSDITFPVSE 289
Db 243 FFREVLLVQKHLDLSNVASYKTYIRELEQSIKAADAVEDLRFRANHGPGMANWQFQE 302
Oy 290 ELEADL-----ASGDQLPMGLVGAQSERFP--SNLEVEASQA 326
Db 303 EWSADLNRLTSRREKKKAVDGVTLTGINTGDQS-----GQNKPGSNLSVPSNP-A 352
Oy 327 SSAEVNASPLNLAHVKNPEQSEEGNVSGHVLGSDVFEP-----MSGMSEAGIPQ 379
Db 353 QSTQLQSS-----YNPFEDEDT-----GSSISEKEDIKAKNVSSYKQTQTYPT 396
Oy 380 SPDDSDSYSGSHSTDSLGMGSP 401
Db 397 DWSDDSENPFSSDTDANGDSNP 418
RESULT 7
CHS5_YEAST
ID CHS5_YEAST STANDARD; PRT; 671 AA.
AC Q12114;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Chitin biosynthesis protein CHS5 (CAL3 protein).
GN CHS5 OR CAL3 OR YLR330W OR L8543.18.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=GRF88;
RX MEDLINE=97265380; PubMed=9111317;
RA Santos B., Duran A., Valdivieso M.H.;
RT "CHS5, a gene involved in chitin synthesis and mating in
RT Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 17:2485-2496(1997).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
FIBRONECTIN TYPE III DOMAIN.
EX MEDLINE=97148176; PubMed=8994808;
RA Bateman A., Chothia C.;
RT "Fibronectin type III domains in yeast detected by a hidden Markov
RT model.";
RL Curr. Biol. 6:1544-1546(1996).
CC -!- FUNCTION: INVOLVED IN CHITIN SYNTHESIS AND ALSO REQUIRED FOR
CC MATING.
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

```









```
QY 373 SEAGIQSPDDSDS 386
Db 317 EPSNIPVSEKDSNS 330

RESULT 11
T2AZ_HUMAN STANDARD; PRT; 1182 AA.
AC Q9Y6Q2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Stoned B-TFIIA-alpha and beta like factor (SALF).
GN SALF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99292779; PubMed=10364255;
RA Upadhyaya A.B., Lee S.H., Dedong J.;
RT Identification of a general transcription factor TFIIAalpha/beta
RT homolog selectively expressed in testis.;
RL J. Biol. Chem. 274:18040-18048(1999).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS, HEART, PLACENTA,
CC KIDNEY, PROSTATE AND UTERUS.
CC -!- SIMILARITY: SOME, IN THE N-TERMINUS TO MEMBERS OF THE ADAPTER
CC COMPLEXES MEDIUM SUBUNITS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF026169; AAD39617.1;
DR HSSP: P32773; 1YTF.
DR MIM: 605357;
DR InterPro: IPR001392; Clathrn_med.
DR InterPro: IPR004855; TFIIA.
DR Pfam: PF00928; Adap_comp_sub; 1.
DR Pfam: PF03153; TFIIA; 1.
KW Transcription regulation; Nuclear protein.
FT DOMAIN 44 150 PRO/SER/THR-RICH.
FT DOMAIN 533 536 POLY-VAL.
SQ SEQUENCE 1182 AA; 132037 MW; 22AD66366D21C371 CRC64;

Query Match 4.9%; Score 106; DB 1; Length 1182;
Best Local Similarity 22.9%; Pred. No.13;
Matches 72; Conservative 42; Mismatches 116; Indels 84; Gaps 15;

QY 128 QIRHSDPESDFVRGK-GEPTVETLSW-HSCRQLLYQAVATILAHAG-----FDCAN 175
Db 750 KVLQSKATEDFFRNSIQSLPFLQLPHSLHQSLQSTASLIVIPAGRTLSFTAEGLGTSN 809
QY 176 ES-----VLETITDVAHEYCKLFTKLLRFVAVDREARLGQTPTFPDYMEQV 219
Db 810 SSANFTFGYPPIHVPAGVLTQVS--GHLYKVNVPIMV--TETSGRAGILQHP--IQOV 862
QY 220 FHEVIGSVLSLQKFWQRIKIDYHSMYLIQSLSEYERYINPEKATEDAKPVKIK--E 277
Db 863 FOOLGQPSVI-----QTSVPQLNPWSLQATTEKSORIETVLQ-QPAILPSGPDVRKHLE 915
QY 278 EPVSDITFPVSE-----LEADLASGQD-----SLPMGVLGQAQSERFPNLEVEASPOASSA 329
Db 916 NATSDILVSPGNKHKIVPALICHQESSHYISLPGVFPQVSQTNDSVSLSGASMA 975
QY 330 EYNASPLNLAHVMEPQSEEGNVSGHGLV-----GSDVFEFPMGSM 372
```

```
Db 976 Q-----NL-----HDESLSTSPHGALHQHVTDIQLHILKRMVGDYVKQPRNIE 1020
QY 373 SEAGIQSPDDSDS 386
Db 1021 EPSNIPVSEKDSNS 1034

RESULT 12
ANK3_HUMAN STANDARD; PRT; 4377 AA.
AC Q12955;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ankyrin 3 (ANK-3) (Ankyrin G).
GN ANK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain stem;
RX MEDLINE=95138209; PubMed=7836469;
RA Kordeli E., Lambert S., Bennett V.;
RT "Ankyrin. A new ankyrin gene with neural-specific isoforms localized
RT at the axonal initial segment and node of Ranvier.";
RL J. Biol. Chem. 270:2352-2359(1995).
CC -!- FUNCTION: Membrane-cytoskeleton linker.
CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
CC alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues.
CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13616; AAA64834.1;
DR HSSP: P55273; 1BI8.
DR Genew: HGNC:494; ANK3.
DR MIM: 600465;
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000488; Death.
DR InterPro: IPR000906; ZU5.
DR Pfam: PF00023; ank; 24.
DR Pfam: PF00531; death; 1.
DR Pfam: PF00791; ZU5; 1.
DR PRINTS: PR01415; ANKYRIN.
DR SMART: SM00248; ANK; 21.
DR SMART: SM00005; DEATH; 1.
DR SMART: SM00218; ZU5; 1.
DR PROSITE: PS50088; ANK_REPEAT; 21.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
FT REPEAT 73 102 ANK 1.
FT REPEAT 106 135 ANK 2.
FT REPEAT 139 168 ANK 3.
FT REPEAT 172 201 ANK 4.
FT REPEAT 203 230 ANK 5.
FT REPEAT 234 263 ANK 6.
FT REPEAT 267 296 ANK 7.
FT REPEAT 300 329 ANK 8.
FT REPEAT 333 362 ANK 9.
FT REPEAT 366 395 ANK 10.
FT REPEAT 399 428 ANK 11.
```

```
FT REPEAT 432 461 ANK 12.
FT REPEAT 465 494 ANK 13.
FT REPEAT 498 527 ANK 14.
FT REPEAT 531 560 ANK 15.
FT REPEAT 564 593 ANK 16.
FT REPEAT 597 626 ANK 17.
FT REPEAT 630 659 ANK 18.
FT REPEAT 663 692 ANK 19.
FT REPEAT 696 725 ANK 20.
FT REPEAT 729 758 ANK 21.
FT REPEAT 762 791 ANK 22.
FT REPEAT 795 825 ANK 23.
FT DOMAIN 1519 1898 SER-RICH.
FT DOMAIN 4090 4174 DEATH.
SQ SEQUENCE 4377 AA; F42379E55768B684 CRC64;

Query Match 4.9%; Score 105.5; DB 1; Length 4377;
Best Local Similarity 18.1%; Pred. No. 81;
Matches 91; Conservative 81; Mismatches 169; Indels 161; Gaps 22;

OY 13 SSSQTNRSSFDLPREFRLVEVDPPLHOPANKPKPTMLDIPSEP-CSLTITHTIOLIQ 71
DB 2311 TSAQHAEDKQMKPLERIIIEVH---TEKGQAEPTEVIIETKKHPEKEMVYQKDLR 2367
OY 72 HNRRLNLI-----ATAQAQNOQTEGVKTESEPLPSCPGSPPLPDDLPLDCKNPN 124
DB 2368 GDINKDLPEKHDADFPCSEGGQOQEEELTAES--LPSY-----LESSRVN 2413
OY 125 APF-QIRHSDPES-----DYRKGGEPTVETLSNHS-----CRQLLYQAVATILA 167
DB 2414 TPVSOEEDSRSSAQLISDDSYK---TLKLLSQHSIEYHDELSLGRGESYFAEKMLL 2469
OY 168 HAGFDCAVESLETLDVAHEYQLKFTLLRFVAVDREAR--LGQTPPDVMEQVFHEV-- 223
DB 2470 SEKLDVSHSDTEESVTHAGPPSELQ-----GSDKRSREKIAATAPKEILSKYKDYSE 2524
OY 224 -GIGSVLSLQKFWHRIKDY-----HSYMLQISK-QLSEVERIVNPKATEDAKPV 273
DB 2525 NGGVKSVDEHFDKVTVLVHSGVNSPKHAMMRFTEDRLDRGREKLIYEDRVDTVKEA 2584
OY 274 KIKEEPVS---DITFPVSELEAD-----LASGD 299
DB 2585 EEKLTVEVSQFTRDKTEKLNDELQPEKKARPKNKEYSSQSPSTSSSPKVLTLLEASND 2644
OY 300 Q-----SLPMG-----VLGAQSERFPNLEVEASQASSA----- 329
DB 2645 EMVKARQHPDQGFPAEKAPSLPSSPEKMWLSQQTEDSKSTVEAKGSISQSKAPDGP 2704
OY 330 -----EVNAPLWNL-----AHVKMEPQSEEGNVSGHGVLGSDVFEPMGSGSEAGIP 378
DB 2705 QSGFQLKQSKLSIRLKFQETHAKSKDMSQEDRKSDQGSRI-----PVKKIQESKLP 2757
OY 379 -----QSPDSDPS 386
DB 2758 VYQVFAREKQKKAIDLDPDESVS 2779

RESULT 13
PAC2_RAT STANDARD; PRT; 488 AA.
AC Q9QY17; Q9QY18; Q9QY19; Q9QY20;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein kinase C and casein kinase substrate in neurons 2 protein
DE (Synaptic dynamin-associated protein II) (Syndapin 2) (Syndapin-II)
DE (SdpII).
GN PACSIN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid-10116;
RN [1]
```

```
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC STRAIN-Sprague-Dawley; TISSUE-Brain;
RX MEDLINE=20171438; PubMed=10704453;
RA Qualmann B., Kelly R.B.;
RT "Syndapin isoforms participate in receptor-mediated endocytosis and
actin organization.";
RL J. Cell Biol. 148:1047-1062(2000).
RN [2]
RP FUNCTION.
RX PubMed=11082044;
RA Modregger J., Ritter B., Witter B., Paulsson M., Plomann M.;
RT "All three PACSIN isoforms bind to endocytic proteins and inhibit
endocytosis.";
RL J. Cell Sci. 113:4511-4521(2000).
CC -!- FUNCTION: May play a role in vesicle formation and transport.
CC -!- SUBUNIT: Homo- and hetero-aggregates with other PACSINS. Binds
dynamitin 1, synaptotagmin, synapsin 1 and the neural Wiskott-Aldrich
syndrome protein (N-WASP).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Vesicle-like cytoplasmic
distribution.
CC -!- ALTERNATIVE PRODUCTS: 4 isoforms; 1/aa (shown here), 2/ab, 3/ba
and 4/bb; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed. Isoforms 1 and 3 are
expressed in PC12 cell line and heart, whereas isoforms 2 and 4
are seen in most tissues examined with higher levels of expression
in muscle, testis and brain.
CC -!- PTM: Phosphorylated by casein kinase 2 (CK2) and protein kinase C
(PKC) (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PACSIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 FCH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF139492; AAF22211.1; -
DR EMBL; AF139493; AAF22212.1; -
DR EMBL; AF139494; AAF22213.1; -
DR EMBL; AF139495; AAF22214.1; -
DR HSSP; P29355; 1SEM.
DR InterPro; IPR001060; Cdc15_Fes_C1P4.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00611; FCH; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50133; FCH; 1.
DR PROSITE; PS50002; SH3; 1.
DR Endocytosis; SH3 domain; Coiled coil; Phosphorylation;
KW Alternative splicing.
FT DOMAIN 11 75 FCH.
FT DOMAIN 428 488 SH3.
FT COILED COIL (POTENTIAL).
FT VARSPPLIC 184 239 MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT VARSPPLIC 302 303 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VARSPPLIC 346 386 MISSING (IN ISOFORM 2 AND ISOFORM 4).
SQ SEQUENCE 488 AA; 55978 MW; B2975012EF0DF56 CRC64;

Query Match 4.8%; Score 105; DB 1; Length 488;
Best Local Similarity 22.5%; Pred. No. 4.5;
Matches 71; Conservative 33; Mismatches 103; Indels 108; Gaps 13;

OY 167 AHAGFDCAVESLETLDV-----AHEYCLKFTLLRFVAVDREA----- 205
DB 134 AEDGFRKQKQWAKKLEVDAAKKAHTACKKEL--AVSREANSKADPSINPQLAKL 190
OY 206 -----RLGQT--PFPPDVMQVFEHVGIGSVLSLQKFWQHRIK 240
```

```
Db 191 ODKIEKQDVLTKDKYKALKELDQTPQYMENMEQVFEQ-----COOFEEKRLR 242
Qy 241 DYHSYMLQISKOLS-----EYERIVNPEKATEDAKPVKIKEEVPSDITRP--V 287
Db 243 FFREVLEVKHLDLSNVASYKGIYRELESGIAKAAVEDLRFWRANHGFCGMANNWPFQ 302
Qy 288 SEELEADL-----ASGDQSLPMGVGLGAQS--ERFPNSLEVEASPOASSAEVNA 333
Db 303 DEWSADLNRTLRSRKKKAADGVTLTGINOTGDSQGNKPSNLSVPSNP-AQSTOLQS 361
Qy 334 SPLWNLAHVMEQSEEGNVSGHVLGSDVFREP-----MSGSEAGIPQSDPDS 386
Db 362 S-----YNPFEDDDT-----GSSVSEKEDIKAKNVSSYKTONYPADWSDDES 405
Qy 387 SYGSHSTDSLWGSSP 401
Db 406 NNFSSIDANGDSNP 420

RESULT 14
Y249_HUMAN STANDARD; PRT; 896 AA.
AC Q2539;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein KIAA0249.
GN KIAA0249
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -1- SIMILARITY: TO HUMAN KIAA0188 AND YEAST SMP2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D87436; BAA13380.1; .
DR Genbank; HGNC:14450; LPIN2.
KW Hypothetical protein.
SQ SEQUENCE 896 AA; 99399 MW; 080113FCCA533272 CRC64;

Query Match 4.8%; Score 105; DB 1; Length 896;
Best Local Similarity 19.1%; Pred. No. 10;
Matches 76; Conservative 56; Mismatches 147; Indels 118; Gaps 16;

Qy 7 WGEIPITSSSTNRSSFDLLPR-----EPRVVEHDPPLHOPSAN----- 45
Db 266 WGGFPSTKVKRSRSDHPTATITPSENTHFRVPSDNLSISEVKEKASMDTIVCTIV 325
Qy 46 KKPPTMLDIPSPCSLTHTIQLIQ---HNRLRLNLIATAQNOQOQTEGVKTESEPL 102
Db 326 KKPRLALGTQMSDPTSA---ELLEPPLESTQISSMLDADHLPNLALEA---PSEK 377
Qy 103 PSCPGSPPL-----PDDLPLDCK--NPNAP---FQIRHSD-----PESDF 138
Db 378 PAAKVDSPSKKGVKHKRSQHQPDDIYLDLKGLEPEVAALVFPKSESEPGSRQWPESDT 437
```

```
Qy 139 YRGKGPVTELSWHSCRQLLYQAVATILAHAGFCANESVLETITDVAHEYC----- 190
Db 438 LSGSQSP-----QVGSAAADSGTECLSDSAMD-LPDVTLSCGLSENSE 482
Qy 191 ---LKFTKTLRFADVAREARGQTPFPDMQOVFHE-----VGIGSVLSLQKFWOHRKDY 242
Db 483 ISKEKFEHIIITTHFAENPGLIDNPMLVIRYRNYNWALAAMPILSLQVFK----- 536
Qy 243 HSYMLQISKOLSEYERIVNPEKA-----TEDAKPVKIKEEVPSDITFPV 287
Db 537 -----SLPKATVESWVKDKMPKKSGRWFWFKRESMTKQLPESKEGKSEAPPASDLPPSS 591
Qy 288 SEELEADLQSGDQSLPMGVGLGAQSERFPNSLEVEASP 324
Db 592 KEPAGARPAENDSSSDEG-----SQLEESITVDPIP 623

RESULT 15
SALL_HUMAN STANDARD; PRT; 1324 AA.
AC Q9NSG2; Q9NSG3; Q9NSG4; Q9P1R0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sal-like protein 1 (Zinc finger protein SALL1) (Spalt-like
DE transcription factor 1) (HSall).
GN SALL1 OR SALL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., DISEASE, AND VARIANTS SER-150 DEL; SER-150 INS AND
RP GLY-159.
RX MEDLINE=99138656; PubMed=9973281;
RA Kohlhasse J., Taschner P.E.M., Burfeind P., Pasche B., Newman B.,
RA Blanka B., Breuning M.H., ten Kate L.P., Maaswinkel-Mooy P.,
RA Mitulla B., Seidel J., Kirkpatrick S.J., Pauli R.M., Wargowski D.S.,
RA Devriendt K., Proesmans W., Gabrielli O., Coppa G.V.,
RA Wesby-van Sway E., Trembath R.C., Schinzel A.A., Reardon W.,
RA Seemanova E., Engel W.;
RT "Molecular analysis of SALL1 mutations in Townes-Brocks Syndrome.";
RL Am. J. Hum. Genet. 64:435-445(1999).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A., DISEASE, AND VARIANTS SER-164 DEL AND
RP GLU-1265.
RX MEDLINE=20004537; PubMed=10533063;
RA Marlin S., Blanchard S., Lacombe D., Denoyelle F., Alessandri J.-L.,
RA Calzolari E., Drouin-Garraud V., Ferraz F.G., Fourmaintraux A.,
RA Philip N., Toubanc J.E., Petit C.;
RT "Townes-Brocks syndrome: detection of a SALL1 mutation hot spot and
RT evidence for a position effect in one patient.";
RL Hum. Mutat. 14:377-386(1999).
RN [3]
RP SEQUENCE OF 26-1324 FROM N.A.
RX MEDLINE=97131507; PubMed=8975705;
RA Kohlhasse J., Schuh R., Dowe G., Kuehnlein R.P., Jaekle H.,
RA Schroeder B., Schulz-Schaeffer W., Kretschmar H.A., Koehler A.,
RA Mueller U., Raab-Vetter M., Burkhardt E., Engel W., Stick R.;
RT "Isolation, characterization, and organ-specific expression of two
RT novel human zinc finger genes related to the Drosophila gene spalt.";
RL Genomics 38:291-298(1996).
RN [4]
RP SEQUENCE OF 313-345 FROM N.A., AND DISEASE.
RX MEDLINE=98085876; PubMed=9425907;
RA Kohlhasse J., Wischermann A., Reichenbach H., Froster U., Engel W.;
RT "Mutations in the SALL1 putative transcription factor gene cause
RT Townes-Brocks syndrome.";
RL Nat. Genet. 18:82-83(1998).
RN [5]
RP DISEASE.
RX MEDLINE=20381976; PubMed=10928856;
```

RA Engels S., Kohlase J., McCaughan J.;

RT "A SALL1 mutation causes a branchio-oto-renal syndrome-like

RL J. Med. Genet. 37:458-460(2000)."

CC -!- FUNCTION: Transcriptional repressor involved in organogenesis (By

CC similarity).

CC -!- SUBUNIT: Interacts with HDAC1, HDAC2, RBBP4, RBP7, MTA1 and MTA2

CC (By similarity). Probably associates with NuRD histone deacetylase

CC complex (HDAC).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- TISSUE SPECIFICITY: Highest levels in kidney. Lower levels in

CC adult brain (enriched in corpus callosum, lower expression in

CC substantia nigra) and liver.

CC -!- DEVELOPMENTAL STAGE: In fetal brain exclusively in neurons of the

CC subependymal region of hypothalamus lateral to the third

CC ventricle.

CC -!- DISEASE: Defects in SALL1 are the cause of Townes-Brocks syndrome

CC (TBS); a rare, autosomal dominant malformation syndrome with a

CC combination of imperforate anus, triphalangeal and supernumerary

CC thumbs, malformed ears and sensorineural hearing loss.

CC -!- DISEASE: Defects in SALL1 are in rare cases associated with

CC hemifacial microsomia (HFM) or Goldenhar syndrome; a common birth

CC defect involving first and second branchial arch derivatives. The

CC highly variable phenotype shows in addition to craniofacial

CC anomalies cardiac, vertebral, and central nervous system defects.

CC -!- DISEASE: Defects in SALL1 are associated with branchio-oto-renal

CC (BOR) syndrome, an autosomal dominant disorder manifested by

CC various combinations of preauricular pits, branchial fistulae or

CC cysts, lacrimal duct stenosis, hearing loss, structural defects of

CC the outer, middle, or inner ear, and renal dysplasia. Associated

CC defects include asthenic habitus, long narrow facies, constricted

CC palate, deep overbite, and myopia. Hearing loss may be due to

CC Mondini type cochlear defect and stapes fixation.

CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER

CC PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Y18265; CAB41400.1; -

DR EMBL; Y18264; CAB41399.1; -

DR EMBL; X98833; CAB41399.1; JOINED.

DR EMBL; AF017655; AAB99908.1; -

DR EMBL; AF074949; AAF19263.1; -

DR HSSP; F15822; IBBO.

DR Genew; HGNC:10524; SALL1.

DR MIM; 602218; -

DR MIM; 107480; -

DR MIM; 164210; -

DR MIM; 113650; -

DR InterPro; IPR000822; znf\_C2H2.

DR Pfam; PF00096; zf-C2H2; 10.

DR Prodom; PD000003; znf-C2H2; 1.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 9.

DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 9.

KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;

KW Zinc-finger; Metal-binding; Repeat; Polymorphism.

FT ZN\_FING 449 471

FT ZN\_FING 477 499

FT ZN\_FING 706 728

FT ZN\_FING 734 756

FT ZN\_FING 766 788

FT ZN\_FING 1001 1023

FT ZN\_FING 1029 1051

FT ZN\_FING 1134 1156

FT ZN\_FING 1162 1184

FT ZN\_FING 150 159

FT DOMAIN 160 163

FT POLY-GLY.

FT	DOMAIN	237	240	POLY-GLN.
FT	DOMAIN	294	297	POLY-ALA.
FT	DOMAIN	371	375	POLY-SER.
FT	DOMAIN	1144	1147	POLY-SER.
FT	VARIANT	150	150	MISSING.
FT	VARIANT	150	150	/FTid=VAR_013155.
FT	VARIANT	150	150	S -> SS.
FT	VARIANT	159	159	/FTid=VAR_013156.
FT	VARIANT	159	159	S -> G.
FT	VARIANT	164	164	/FTid=VAR_013157.
FT	VARIANT	164	164	MISSING.
FT	VARIANT	1265	1265	/FTid=VAR_013158.
FT	VARIANT	1265	1265	G -> E.
FT	VARIANT	79	79	/FTid=VAR_013159.
FT	CONFLICT	79	79	G -> A (IN REF. 1; CAB41399).
SQ	SEQUENCE	1324 AA;	140390 MW;	CBE787847A96D8AA CRC64;

Query Match 4.8%; Score 103.5; DB 1; Length 1324;

Best Local Similarity 22.0%; Pred. No. 22;

Matches 87; Conservative 53; Mismatches 140; Indels 115; Gaps 20;

QY 34 VHDPLHOPSAKPKPTMLDIPSEPCSLTHTI-----OLIQHN--RRLRNLIAQA 85

DB 74 VNEPGSPETFSPPP--DNPDEQMDTVNKTDQVDCSLSEHGLDRESMEVAPV 131

QY 86 QNQOQTEGVKTESEPLPSCPGSPPLPDDLPLDCKNPNAPFQIRHSDPESDFYRGKEP 145

DB 132 ANK---SGSGTSSGSHSSTAPSS-----SSSSSSSSSGGGSS 165

QY 146 VTELSWHSCROLLYQAVATILAHAG--FDCANESVLETLTDVAHEYCLKFKTLRFAVDR 203

DB 166 STGTS-----AITSPLQLGLTLTLGNFVINSNVIIEN---LQSTKVAQAQFSQ 212

QY 204 EARL-----GOTPPDVMEQVFEVHGIGSVLSLQKFWQHRKIKQVHSMQLQSKOLSEYE 258

DB 213 EARGGASGKLAVPALMEQ-----LLALQ---QQQI-----HQLQIEIRHQIL 255

QY 259 RIVN-----PEKATEDAKPVKKEEVPVSDITFPVSEELADLASC-DQSLPMGVLAGQSE 312

DB 256 LLASONADLPSTSSPSQGTLRSTANPLSTLSSHLSQLAA--AAGLAQSL----- 303

QY 313 RFPNLEVEASPOASSAEVYNASPLWNLAHVMEQSEEGNMGVGLGSDVFPEPMSCM 372

DB 304 -----ASQASASISGVKQLPPIQL-----PQSS-----SGNTIIPNSGSSPNMNI 343

QY 373 SEAGI--PQSPDDSDSVGSHSTDSLM--GSSPVF 403

DB 344 LAAAVTTPSEKVAAGASGASHVSNPVSSTSSSPAF 378

Search completed: July 18, 2003, 09:57:50

Job time : 31.5071 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 09:48:28 ; Search time 45.128 Seconds  
(without alignments)  
881.929 Million cell updates/sec

Title: US-09-857-308-1

Perfect score: 2175

Sequence: 1 MNLQRYWGEIPISSQTNRS.....SLMGSPVFNQCKRMKRI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142.5	6.6	979	T40006	hypothetical prote
2	126	5.8	1621	A82255	hypothetical prote
3	123	5.7	552	A31401	macrophage colony-
4	123	5.7	552	A31401	colony-stimulating
5	121.5	5.6	2957	S35703	hypothetical prote
6	117	5.4	575	A69716	spore coat assembl
7	112.5	5.2	1332	S41552	probable transcrip
8	112	5.1	609	T28736	hypothetical prote
9	110.5	5.1	785	S54016	SOK2 protein - yea
10	108.5	5.0	671	S53407	CHS5 protein - yea
11	108.5	5.0	1206	T34021	protein kinase SK2
12	107	4.9	2484	T26216	hypothetical prote
13	107	4.9	2607	T26215	hypothetical prote
14	106	4.9	387	A43704	involucrin - weste
15	106	4.9	1180	T20773	hypothetical prote
16	106	4.9	4549	T20771	hypothetical prote
17	106	4.9	4667	T20774	hypothetical prote
18	106	4.9	5170	T15348	hypothetical prote
19	105.5	4.9	444	T25111	hypothetical prote
20	105.5	4.9	4377	A55575	ankyrin 3, long sp
21	105	4.8	1163	A36685	205K microtubule-a
22	104.5	4.8	727	A88131	protein F10G7.9 [l
23	104.5	4.8	834	D57282	ankyrin-related pr
24	104	4.8	1186	D72229	meiotic recombinat
25	103.5	4.8	914	T21122	hypothetical prote
26	103.5	4.8	2224	1 KFHU5	coagulation factor
27	103	4.7	1403	1 A47328	natural killer cel
28	103	4.7	1826	D86502	excinuclease ABC s
29	102.5	4.7	679	T24365	hypothetical prote

#### RESULT 1

T40006

hypothetical protein SPBC25H2.11lc - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000

C;Accession: T40006

R;Durso, G.; Lye, G.; Bowman, S.; Church, C.; Wood, V.; Barrell, B.G.; Rajandream, M.

submitted to the EMBL Data Library, April 1997

A;Reference number: Z21898

A;Accession: T40006

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-979 <DUR>

A;Cross-references: EMBL:Z95397; PIDN:CAB08777.1; GSPDB:GN00067; SPDB:SPBC25H2.11lc

A;Experimental source: strain 972h; cosmid c25H2

C;Genetics:

A;Gene: SPDB:SPBC25H2.11lc

A;Map position: 2

A;Introns: 192/1; 258/3; 292/3; 339/1; 753/3; 801/3; 873/2

A;Superfamily: bromodomain homology

F;324-379/Domain: bromodomain homology <BRO>

Query Match 6.6%; Score 142.5; DB 2; Length 979;

Best Local Similarity 19.4%; Pred. No. 0.024;

Matches 83; Conservative 74; Mismatches 183; Indels 87; Gaps 14;

QY 10 IPISSQT--NRSSFDLLPREFLVEVHDPPLHQPSANKPKPTMLDIPSEPCSLFIHTI 67

Db 596 VPMTAGSVLETESEEDLYPRDYSLFEI-----NRNTPGVF-----SLMYKNI 637

QY 68 QLIQHNRRLNLIAQAQNOQTEGVTKESEPLPSCPGSPPLPDDLLPLDCKNNAFF 127

Db 638 AKMQEIRKLCNKIQTVRQLQLPFPYIEHKKSH-VPFANNEPILDD--IPQYNNMSPK 694

QY 128 QRRHSDPESDFYRGKGPVTELSWHSRCLLLQAVATILAHAGFDCANESVLETLDVAH 187

Db 695 PLAH-----DVLKKLCTLIILFHAGFESFGWALDLEIAA 730

QY 188 EYCLK-----FTKLLRFVDRARLGTQFPDVMQVFEHVGIGSVLSQKFWQHRKIDVH 243

Db 731 DYMAKMGAVMDQVLIYKDKSQO-----EIVGQTLGELGVDVNDLISYVYHVERQS 783

QY 244 SYMLQISKOLSEYERIVNPEKATEDAKPVKKEEVPVSDITFPVSE-----L 291

Db 784 VKLLEIHLQHRHFVLELLRPAISERDEAIFNQNGESFVTGNFYSYETGDDFFGLRELGL 843

QY 292 EADLASGDQSLPMGVLAGQSERFPPSNLEVEASPOAS--SAEVNASPLWNLAHVKMEPQSE 350

Db 844 DRELGLDLSVPLHLL---QSLRNSNSWQ--PEATIKGDOEYAPPKYPPPTAESISNE 898

QY 351 EGNVSGHGVIGSDVF-----EPPMSGSEAGIPQSDSDSSYSGHSHSTSLMGSSPVF 403

Db 351 EGNVSGHGVIGSDVF-----EPPMSGSEAGIPQSDSDSSYSGHSHSTSLMGSSPVF 403



Db 485 -SSDQIPESVFHLLVPGIILVLLTVGGLLFYKWKWRSHRDPOTLDSVSGRPESSSL 540

RESULT 4

S35703

colony-stimulating factor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S35703

R:Borycki, A.G.; Lenormand, J.L.; Guillier, M.; Leibovitch, S.A.

Biochim. Biophys. Acta 1174, 143-152, 1993

A:Title: Isolation and characterization of a cDNA clone encoding for rat CSF-1 gene. Pos

A:Reference number: S35703; MUID:93363632; PMID:8357831

A:Accession: S35703

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-552 <BOR>

A:Cross-references: EMBL:M84361; NID:g203640; PIDN:AAA03032.1; PID:g203641

C:Superfamily: macrophage colony-stimulating factor

Query Match 5.7%; Score 123; DB 1; Length 552;

Best Local Similarity 21.6%; Pred. No. 0.29;

Matches 90; Conservative 47; Mismatches 144; Indels 136; Gaps 20;

QY 14 SSGTNRSSDPLPREPLVEVHDPPHQPANKPKPTMLDIPSEPCSLTIHTIQLIQRN 73

Db 226 SORTEGSS--LLPSELPL-----RIEDPGSAKQRP-----PRSTCQ-TLESTEQNHG 270

QY 74 RLRLNLIAQAQNOOQTGKVTSESEPLPFCGSPPLPDDLLPLDCKNPNAPFQIRHSD 133

Db 271 DRL-----TEDSQHPGAGGVPVGVEDIL----- 294

QY 134 PESDFYRGKGPVTELSWHSRQLLYQAVATILAHAGFCANESVLETLTDAHEYCLKF 193

Db 295 -ESSL--GYNWLEASGASEGFLTQ-----EAKESPSTPVGGS----- 331

QY 194 TKLLRAVDREARLGOTPPP-----DVMQVEFHEVG-----IGSVLSLQKFWOHRIK 240

Db 332 ---IQAETDRPRALSAPFPKSTEDQKVPDITDRPLTEVNPMPRIQOT-----QNNTP 381

QY 241 DYHSYMLQISKOLSEYERIVNPEKATEDAKPVKIKEEVPDITFPVSELEADLASGQ 300

Db 382 E----KTDGTSTLRDQHGSPHIATPNPQVRSNATPVAQLLLPKSH-----SWG 430

QY 301 SLPMGVL----GAQSRFPFNSLEVEASPOASS---AEVNASPLNLAHVKMEPQSEEGN 353

Db 431 VLPLGELCKRSTRDRSRFAELGGSGASEGARPVAFNSIPLTDGTHV-----EQHEG- 484

QY 354 VSGHGVLSGVDFEPPMSG-----MSEAGI-----PQSPDDSDSSYSGSHSTDLSL 396

Db 485 -SSDQIPESVFHLLVPGIILVLLTVGGLLFYKWKWRSHRDPOTLDSVSGRPESSSL 540

RESULT 5

hypothetical protein T04D1.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T33152

R:Davidson, S.; Wohldmann, P.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid T04D1.

A:Reference number: Z21292

A:Accession: T33152

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2957 <DAY>

A:Cross-references: EMBL:AF067617; PIDN:AAC17559.1; GSPDB:GN00019; CESP:T04D1.4

A:Experimental source: strain Bristol N2; clone T04D1

C:Genetics:

A:Gene: CESP-T04D1.4

A:Map position: 1

A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3

Query Match 5.6%; Score 121.5; DB 2; Length 2957;

Best Local Similarity 20.0%; Pred. No. 4.5;

Matches 86; Conservative 44; Mismatches 172; Indels 127; Gaps 17;

QY 37 PPLHQPSANKPK-----PPTMLDIPSEPCSLTIHTIQLIQRNRLNLATAQ 84

Db 132 PPPHYSSNGSPGQGYRPGFVTNGPPPGQGYPPP--QAPQNYQQMYHOOR-----AQOQ 184

QY 85 AQNOOQTGKVTSESEPLPSCP-----GSPPLPDDLPLDCKNPNAPFQIRHSDPES---- 136

Db 185 QAGCQQAQG--PPGYGPGYAPQASTGYPPPPSQQSPY-APQAGPPQMRHQYPPHSQQQ 241

QY 137 -----DFYRGKGPVTELSWHSRQLLYQAVATILAHAGFCANESVLETLTDAHEYC 190

Db 242 APPGYWDGYGYPGPPPS-----QQCQQQGGGVPVTAPOSMAOQ----- 281

QY 191 LKFTKLLRAVDREARLGOTPPFDVMEQVFH-----EVGIGSVLSLQKF-----WQ 236

Db 282 -----EQWGRVGTNDLMNVIAQVEKSAECKARIQQLYNNMORSASADAQVQQLQ 330

QY 237 HRIKDYHSMLOISKOLSB-EYERIVNPEKATEDAKPVKIKEEVPDITFPVSELEADL 295

Db 331 HKLAQFQSDHYRYSQQLQHVQMQQQQQQAAAA-----AAAAAAS 373

QY 296 ASGQDQLPMGLGAQSERFPSN---LEVEASPOASSAEVNASPLNLAHVKMEPQSEEG 352

Db 374 AGGHGQPPPTTSQSNQIPSPNSVPVHVNQPSQVHITADKTRTLISVYHD----- 426

QY 353 NVSGHGVLSGSD-----VFPEPMGSMSEAGIQSPD-----DSDSSYSGSHSTDLSM 397

Db 427 ---GYGAGSGAGSSTESKESIVPPEPTPSNQESFPVNPAPQVKMEPKSPYQQQYNGMEN 483

QY 398 GSSPVENOR 406

Db 484 GSGNPTAAQ 492

RESULT 6

A69716

spore coat assembly protein (spovID) - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C:Accession: A69716; B47083

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Ber

C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tojnoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: A69716

A:Status: nucleic acid

A:Molecule type: DNA

A:Residues: 1-575 <KUN>

A:Cross-references: GB:Z99118; GB:AL009126; NID:g2635200; PIDN:CAB14771.1; PID:el1840

A:Experimental source: strain 168

R:Beall, B.; Driks, A.; Losick, R.; Moran Jr., C.P.

J. Bacteriol. 175, 1705-1716, 1993

A:Title: Cloning and characterization of a gene required for assembly of the Bacillus

A:Reference number: A47083; MUID:93194796; PMID:8449878

A:Accession: B47083

A:Status: preliminary

A:Molecule type: nucleic acid









Db 552 ASPRH-----QEE-----DIELEAPKDATSVAVEP--SNEDVKPEEKGEAEADDI 596  
Qy 352 GNVSGHGVLDVPEPMSGSEAGIPQSPDDSDSYGSHSTDSL-----GSSPV 402  
Db 597 NNVSKAASGESTHOKTEASASLESSAVTEEQETTEAEVNDVLDVLTKEAKKNTGNSN 656  
Qy 403 FNORCKKMRK 413  
Db 657 NKKNKKKKKK 667  
RESULT 11  
T34021  
protein kinase SK2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34021  
R:Fukami, Y.; Yamamoto, H.; Ichihara, T.; Mori, K.; Gomi, T.; Sato, K.  
submitted to the EMBL Data Library, April 1997  
A:Description: SK2, a putative rat homologue of yeast protein kinase NRK1.  
A:Reference number: 221463  
A:Accession: T34021  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1206 <FUK>  
A:Cross-references: EMBL:AB003357; NID:d1106513; PID:d1020890; PIDN:BA20077.1  
A:Experimental source: strain SD  
C:Genetics:  
A:Gene: SK2

Query Match 5.0%; Score 108.5; DB 2; Length 1206;  
Best Local Similarity 20.3%; Pred. No. 11;  
Matches 128; Conservative 68; Mismatches 180; Indels 255; Gaps 31;  
Qy 6 YWGEIPISSSTNRS-----SFDLLPREFRLEVEVH--DPPHQPSS-----ANKPKPPT 51  
Db 195 YWMAPEVVMCTSKDRPYDKADVMSGLITLIEMAEIPEPHHLPNMRVLLKIAKSEPT 254  
Qy 52 MLDIPSPCSLTIH-----TIQLTOH-----NRRLRNLATAQAQ 86  
Db 255 L-----AQSRWSSNFKDPLKCLBKNDVARTTSLLQHPVTVDSNKNPVRELLAEAKA 310  
Qy 87 NQOQTEGVK-----TESEPLP-----TTESEPLP-----103  
Db 311 VTEEDGKEDEDDDETESALPIANKRASSDLSIASEEDKLSQACILLESVSERTEHN 370  
Qy 104 -----SCPGSPPLPDLLPLD--CKNPNAPFOIRHSDPESDFY---RGKGPEVT 147  
Db 371 TSGDKFSNKVLSEKPTPEGPEKTVDDGPANDVNLETVAEPNDQAVGFHENGREKKRPQL 430  
Qy 148 ELSWHSRCQLLYQAVATILAHAGFDCANESVLETLTDVAHEYCLKFTKLLRFAYDR--- 203  
Db 431 ESQPTDEQ---QTVDNVLVGEQND--SNIVILETND-----CLKPEE-----DRNEEN 475  
Qy 204 ----EARLQGT--PPFDV-----MEQVFHEVGIGSVLSLQKFWQHRK 240  
Db 476 QEIIENKLQSEEEKDHIQTMDLVSQETGEKADFQALDNEVGFTEETQEKLG----K 531  
Qy 241 DYHMYLQISKOLSE-----EYERIVNP-----EKATEDAKP 272  
Db 532 DDKTHKVVISDITSEVGTDEPPGDTOKSAEQSDAEGGAGEAPEAQTUTEKATEGPEA 591  
Qy 273 VKIKEEP-----VSDITFPVSEEL---EADLASGDQSLP 303  
Db 592 HGAEERPSGRVEDKQEQSVAECGEGQVTSSESTRATTEETPEDEVDQVSESNSIE 651  
Qy 304 ----MGVLGAOSERFSS-----NLEVASPQASSAENVASPLNLAHVKNMEPO----- 347  
Db 652 ELERLGTGAEEQALGKGAATELDLERENAEQELPVKAEPQAPAAQSEAPPPVLIPS 711  
Qy 348 ---RSEGNVSGHGVLDVPEPMSGSEAGIPQSPD-----DSDSSYGS---HSTDLSL 396  
Db 195 YWMAPEVVMCTSKDRPYDKADVMSGLITLIEMAEIPEPHHLPNMRVLLKIAKSEPT 254

Db 712 INIHSE--NTENKGMGA--LPKP-----ETILPEPENGKGNDDTSGTGSTVENSSDL 762  
Qy 397 MGSSPVF-----NORCKKMRK 413  
Db 763 NLSISSELSKTKDSCSVSLQETRRQKTKLKK 793  
RESULT 12  
T26216  
hypothetical protein W06A7.3c - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26216  
R:Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: 220173  
A:Accession: T26216  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2484 <WIL>  
A:Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GNO0023; CESP:W06A7.3c  
A:Experimental source: clone W06A7  
C:Genetics:  
A:Gene: CESP:W06A7.3c  
A:Map position: 5  
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match 4.9%; Score 107; DB 2; Length 2484;  
Best Local Similarity 19.2%; Pred. No. 42;  
Matches 85; Conservative 70; Mismatches 168; Indels 120; Gaps 20;  
Qy 28 EFRLEVEVDHPPHQPSSANKPKPTMLDIPSPCSLTIHTIQLIQHNRRRLRLIATAQAQ 87  
Db 1317 EQSIANWIDELVHEDDEKK-----VPEVTANISVSASENID-DSTANAVKTEVSS 1367  
Qy 88 QQ-----QTEGVKTEES-----EPL-----PSCPGSPPLPDLLPL----- 118  
Db 1368 EQLQVATVFELESAPESAEESAAIPEVQEPLEKVEVQPDLSQNSAPHKIIDLHFNIPKDH 1427  
Qy 119 -DCKNPNAPFOIRHSDPESDFYCKGE-----PVTELSWHSRCQ-----LLYQAVATI 165  
Db 1428 EDYGNQVVPFETESSE--ESQKADGNQENQEEEDVVAELNHFPIRQWRDEDIVISLQSLKSL 1486  
Qy 166 LAHAGFDCANESVLETLTDVAHEYCLKFTK-----LLRFADVREARLGOTPPFDVMEQ 218  
Db 1487 VAEVG--CITDVSADVNEQDEESTLKLKVPSEPSLELDFDNDPKVIHVPIP-LMEP 1543  
Qy 219 VFHEVGIGSVLSLQKFWOHRKIDYHMYLQISKOLSEYERIVNPERKATEDAKPVKTK-- 276  
Db 1544 -----ATMYLEEMVEWIIAD-----AVKEVSE--MEVVTSEISEMAPQVSESTC 1586  
Qy 277 --EPPVSDITFPVSEELAEADLASGDQSLPMGVLGAQSERPSPNLEVASP----- 324  
Db 1587 PIPEPLADLKLUPVDEDEKT-----PEPEVPVPGVQOERIPIPEVQOAPITPQRPAP 1639  
Qy 325 -----QASSAENVASPLNLAHVKNMEPOEESGVNHSGLVGLSGDVFEPEPMSGMS 373  
Db 1640 KSELPKVAKPLDDSKSRVREAPL-----NIKLGRTYSEEQKE-----LVESLERPLTIT 1690  
Qy 374 EAGIPQSPDDSDSYGSHSTDSL 396  
Db 1691 QOKPPEKPTEDIGALSPLSPNTL 1713  
RESULT 13  
T26215  
hypothetical protein W06A7.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26215  
R:Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: 220173

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2003, 07:48:47 ; Search time 91.237 Seconds  
(without alignments)  
604.643 Million cell updates/sec

Title: US-09-857-308-1  
Perfect score: 2175  
Sequence: 1 MNLQRYWGEIPISSQTNR.....SLMGSPVFNQCKRMKRI 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2175	100.0	414	21	Human ART-1 protei
2	2175	100.0	419	22	Human protein sequ
3	2155	99.1	412	22	Human cell death p
4	2121	97.5	403	21	Gene 41 human secr
5	2059.5	94.7	404	21	Human secreted pro
6	1898	87.3	363	22	Human cell death p
7	1022	47.0	199	22	Human cell death p
8	915	42.1	198	20	Human endometrium
9	858	39.4	169	22	Human cell death p
10	572	26.3	111	22	Human cell death p

11	371	17.1	71	22	AAG98658	Human cell death p
12	239	11.0	46	22	AAG98662	Human cell death p
13	224	10.3	43	22	AAG98663	Human cell death p
14	132	6.1	359	22	ABE62430	Drosophila melanog
15	131	6.0	1049	22	ABE58717	Drosophila melanog
16	123	5.7	520	22	AAE00578	Murine colony stim
17	123	5.7	520	22	AAE00579	Murine colony stim
18	123	5.7	520	22	AAE00580	Murine colony stim
19	123	5.7	552	9	AAE03390	Deduced sequence f
20	123	5.7	552	17	AAW10070	Murine long form C
21	123	5.7	552	18	AAW36145	Mouse long form of
22	123	5.7	552	18	AAW35752	Murine colony stim
23	123	5.7	552	18	AAW22615	Murine 4 kb colony
24	123	5.7	552	21	AAE19547	Human colony stimu
25	123	5.7	552	21	AAE90329	Mouse CSF-1 protei
26	123	5.7	552	22	AAE00540	Murine colony stim
27	123	5.7	552	22	AAE59606	Murine colony stim
28	123	5.7	552	22	AAE49459	Murine colony stim
29	123	5.7	552	23	ABE57255	Murine CSF-1 #1.
30	121.5	5.6	2977	22	ABE69480	Mouse ischaemic co
31	118	5.4	377	17	AAW05411	Drosophila melanog
32	117	5.4	486	21	AAE43338	Human H74 protein.
33	117	5.4	486	21	AAE54041	Human ORFX ORF3102
34	115	5.3	783	23	AAE82709	Protein encoded by
35	115	5.3	1030	22	ABE23699	Amino acid sequenc
36	114	5.2	552	9	AAE80766	Novel human diagno
37	114	5.2	552	17	AAW10071	Deduced sequence f
38	114	5.2	552	18	AAW36146	Murine short form
39	114	5.2	552	18	AAW35753	Mouse short form o
40	114	5.2	552	18	AAW22616	Murine colony stim
41	114	5.2	552	21	AAE19548	Murine 2 kb colony
42	114	5.2	552	21	AAE90330	Human colony stimu
43	114	5.2	552	21	AAE90330	Mouse CSF-1 protei
44	114	5.2	552	22	AAE00541	Murine colony stim
45	114	5.2	552	22	AAE59607	Murine colony stim
			552	22	AAE49460	Murine CSF-1 #2.

ALIGNMENTS

RESULT 1

AAE03880

ID AAB03880 standard; protein; 414 AA.

XX

AC AAB03880;

XX

DT 26-OCT-2000 (first entry)

XX

DE Human ART-1 protein sequence.

XX

DE Human: tumour antigen protein; ART-1; HLA antigen; cytostatic;

KW bound cytotoxic T cell; tumour; cancer; remedy; diagnose.

XX

OS Homo sapiens.

XX

PN WO2000032770-A1.

XX

PD 08-JUN-2000.

XX

PF 30-NOV-1999; 99WO-JP06682.

XX

PR 01-DEC-1998; 98JP-0341253.

XX

PA (SUMU ) SUMITOMO PHARM CO LTD.

PA (ITOH/) ITOH K.

XX

PI Itoh K, Gomi S;

XX

DR WPI: 2000-412318/35.

DR N-PSDB; AAA62864.

XX

PT Novel tumor antigen protein ART-1, tumor antigen peptide originating from it, their derivatives, and DNAs, applicable in vivo or in vitro as

PT remedies, preventives and diagnostics for tumors -

PS Claim 1; Page 44-46; 59pp; Japanese.

XX The invention relates to a novel human tumour antigen protein, ART-1.  
CC Included in the invention are polynucleotide sequences encoding the ART-1  
CC protein, and mutated ART-1 proteins which when broken down  
CC intracellularly produce a tumour antigen peptide that can recognise HLA  
CC antigen and bound cytotoxic T cells. Antibodies which specifically  
CC recognise ART-1 and its derivative peptides, are also included in the  
CC invention. ART-1 exhibits cytostatic activity. The tumour antigen  
CC protein, tumour antigen peptide originating from it, their derivatives,  
CC and DNAs are applicable in vivo or in vitro as remedies, preventives and  
CC diagnostics for tumours.  
XX The present sequence represents the human ART-1 protein sequence.

XX Sequence 414 AA;

Query Match 100.0%; Score 2175; DB 21; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.4e-191;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSSANKPKPPTMLDIPSEPC 60  
DB 1 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSSANKPKPPTMLDIPSEPC 60

QY 61 SLTIHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLLPDC 120  
DB 61 SLTIHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLLPDC 120

QY 121 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 180  
DB 121 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 180

QY 181 TLTDVAHEYCYLKFTKLLRFVAVDRARLGQTFFPDVMEQVFEHVGIGSVLSQKFWOHRK 240  
DB 181 TLTDVAHEYCYLKFTKLLRFVAVDRARLGQTFFPDVMEQVFEHVGIGSVLSQKFWOHRK 240

QY 241 DYHSMYLIQSKQSEERIVNPEKATEDAPVKIKEPVSDFIYFVSEELADLASGDQ 300  
DB 241 DYHSMYLIQSKQSEERIVNPEKATEDAPVKIKEPVSDFIYFVSEELADLASGDQ 300

QY 301 SLPMGVLAQSERPPSNLEVEASPOASSAEVNASPLNLAHVMEQVFEHVGIGSVLSQKFWOHRK 360  
DB 301 SLPMGVLAQSERPPSNLEVEASPOASSAEVNASPLNLAHVMEQVFEHVGIGSVLSQKFWOHRK 360

QY 361 GSDVFEPMGMSAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 414  
DB 361 GSDVFEPMGMSAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 414

RESULT 2

AAM25845  
ID AAM25845 standard; Protein; 419 AA.

XX AAM25845;

XX 16-OCT-2001 (first entry)

XX Human protein sequence SEQ ID NO:1360.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnary; antiulcer; osteopathic; eczema;  
KW dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic;  
KW immunoprotective; antineoplastic; nontropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;

KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

XX Homo sapiens.

XX WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-457603/49.

XX N-PSDB; AAH99786.

XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX Claim 20; Page 281; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX AAM25963. The proteins can have activities based on the tissues and  
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;  
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX cardiovascular; antianaemic; antiagregant; haemostatic; vulnary;  
XX antiulcer; osteopathic; dermatological; antiallergic; antisthmatic;  
XX antidiabetic; cytostatic; neuroprotective; antineoplastic; nontropic;  
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX encoding them can be used in gene therapy, antisense therapy and vaccine  
XX production. The proteins and polynucleotides are useful for screening for  
XX agonists or antagonists of a protein and for the treatment and diagnosis  
XX of disorders associated with the activity of a protein e.g. inflammation,  
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,  
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
XX osteoporosis, severe combined immunodeficiency, eczema, allergic  
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and  
XX neurological disorders.

XX Sequence 419 AA;

Query Match 100.0%; Score 2175; DB 22; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.4e-191;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSSANKPKPPTMLDIPSEPC 60

DB 6 MNLQRYWGEIPISSQTNRRSFDLLPREFRLVEVDHPPHQPSSANKPKPPTMLDIPSEPC 65

QY 61 SLTIHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLLPDC 120

DB 66 SLTIHTIQLIOHNRRLNLIATAQAOQOQTEGVKTESEPLPSCPGSPPLPDDLLPDC 125

QY 121 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 180

DB 126 KNPAPFOIRHSDPESDFYRGKGPEVTELSWHSCRQLLYQAVATILAHAGFDCANESVLE 185

QY 181 TLTDVAHEYCYLKFTKLLRFVAVDRARLGQTFFPDVMEQVFEHVGIGSVLSQKFWOHRK 240

DB 186 TLTDVAHEYCYLKFTKLLRFVAVDRARLGQTFFPDVMEQVFEHVGIGSVLSQKFWOHRK 245

QY 241 DYHSMYLIQSKQSEERIVNPEKATEDAPVKIKEPVSDFIYFVSEELADLASGDQ 300

Db 246 DYHSYMLQISKQLSEYERIVNPEKATEDAKPVYKIKEEPVSDITFPVSELEADLSDGQ 305  
 QY 301 SLPMGVLGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKNPEQSEEGNVSGHVL 360  
 Db 306 SLPMGVLGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKNPEQSEEGNVSGHVL 365  
 QY 361 GSDVFEPMGSMSEAGIPQSPDDSSYSGSHSDSLMGLSSPVFNQCKKRMRI 414  
 Db 366 GSDVFEPMGSMSEAGIPQSPDDSSYSGSHSDSLMGLSSPVFNQCKKRMRI 419  
 RESULT 3  
 ID AAG98644  
 XX AAG98644 standard; Protein; 412 AA.  
 AC AAG98644;  
 XX  
 DT 21-SEP-2001 (first entry)  
 DE Human cell death protective cDNA clone CNI-00714 ORF1 protein, SEQ:77.  
 DE Cell death protective; apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
 KW stroke; cerebral infarction; ischaemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.  
 OS Homo sapiens.  
 OS  
 PN WO200145638-A2.  
 PN  
 XX 28-JUN-2001.  
 PD  
 XX 11-DEC-2000; 2000WO-US33547.  
 PF  
 XX 14-DEC-1999; 99US-0461697.  
 PR  
 XX (COGE-) COGENT NEUROSCIENCE INC.  
 PA  
 XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
 PI WPI: 2001-390297/41.  
 XX N-PSDB; AAH84170, AAH84171.  
 DR  
 XX Novel protective sequence polynucleotides and polypeptides, used to  
 PT identify modulators of their expression and activity, which are used in  
 PT to treat central nervous system conditions, diseases and disorders -  
 XX  
 PS Claim 1; Fig 6A; 325pp; English.  
 XX  
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
 CC while the remaining nucleic acid sequences within the range given above  
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
 CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
 CC protective ORFs. The cell death protective cDNA clones are able to  
 CC prevent, delay or reverse progression through the apoptotic or necrotic  
 CC pathways when injected into a cell predisposed to or undergoing cell  
 CC death. The cell death protective nucleic acids and polypeptides can be  
 CC used in the diagnosis and treatment of disorders associated with cell

CC death, and to screen for compounds which modulate their activity or  
 CC expression. Such modulators, preferably a small organic molecule, an  
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
 CC cell death-related diseases. Such diseases include those associated with  
 CC the central nervous system including psychiatric or neurological  
 CC disorders, especially ischaemia-related conditions such as strokes, and  
 CC also includes neurodegenerative disorders such as Alzheimer's disease,  
 CC Huntington's disease, or Parkinson's disease. The modulators may also be  
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
 CC vascular diseases such as ischaemic encephalopathy or cerebral  
 CC infarction; eye conditions such as diabetic retinopathy or macular  
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
 CC respiratory conditions such as asthma or chronic obstructive pulmonary  
 CC disease; neoplastic conditions such as cancers or benign tumours; blood  
 CC cell conditions such as anaemia; gastrointestinal conditions such as  
 CC gastritis or ulcerative colitis; liver conditions such as biliary  
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
 CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.  
 XX  
 SQ Sequence 412 AA;  
 Query Match 99.1%; Score 2155; DB 22; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-190;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 RYWGEPISSTQNRSSFDLLPREFRLVEVHPDPLHQPSSANKPKPTMLDIPSECSLTI 64  
 Db 3 RYWGEPISSTQNRSSFDLLPREFRLVEVHPDPLHQPSSANKPKPTMLDIPSECSLTI 62  
 QY 65 HTIQLIHNRLRLNIATAQAQQQTEGVKTESEPLSCPGSPPLPDDLLPDKCKNPN 124  
 Db 63 HTIQLIHNRLRLNIATAQAQQQTEGVKTESEPLSCPGSPPLPDDLLPDKCKNPN 122  
 QY 125 APFQIRHSDPESDFYRGKGPVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 184  
 Db 123 APFQIRHSDPESDFYRGKGPVTELSWHSCROLLYQAVATILAHAGFDCANESVLETLTD 182  
 QY 185 VAHEYCLKFTKLLRFPAVDREARLGOTPPFDVMEQVFHEVGIGSVLSLQKFWQHRDKYHS 244  
 Db 183 VAHEYCLKFTKLLRFPAVDREARLGOTPPFDVMEQVFHEVGIGSVLSLQKFWQHRDKYHS 242  
 QY 245 YMLQISKQLSEYERIVNPEKATEDAKPVYKIKEEPVSDITFPVSELEADLSDGQSLPM 304  
 Db 243 YMLQISKQLSEYERIVNPEKATEDAKPVYKIKEEPVSDITFPVSELEADLSDGQSLPM 302  
 QY 305 GVLGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKNPEQSEEGNVSGHVLGSDV 364  
 Db 303 GVLGAQSERFPNLEVEASPOASSAEVNASPLWNLAHVKNPEQSEEGNVSGHVLGSDV 362  
 QY 365 FEPMGSMSEAGIPQSPDDSSYSGSHSDSLMGLSSPVFNQCKKRMRI 414  
 Db 363 FEPMGSMSEAGIPQSPDDSSYSGSHSDSLMGLSSPVFNQCKKRMRI 412  
 RESULT 4  
 AAB45216  
 ID AAB45216 standard; Protein; 403 AA.  
 XX  
 AC AAB45216;  
 XX  
 DT 12-FEB-2001 (first entry)  
 DE Gene 41 human secreted protein homologous amino acid sequence #157.  
 DE  
 XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;  
 KW ophthalmological; autoimmune disease; hyperproliferative disorder;

KW cardiovascular disorder; cerebrovascular disorder; wound healing;  
KW nervous system disorder; aging; chemotaxis.  
OS Homo sapiens.  
XX WO200058467-A1.  
XX 05-OCT-2000.  
XX 22-MAR-2000; 2000WO-US07505.  
XX 26-MAR-1999; 99US-0126502.  
XX 17-DEC-1999; 99US-0172410.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI; 2000-611712/58.  
XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
PS Disclosure; Page 60-61; 440pp; English.  
XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human  
CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent  
CC alternative polypeptides encoded by the genes, and amino acid sequences  
CC to which they are homologous. The genes and proteins have activities  
CC dependent on the tissues and cells in which they are expressed. Examples  
CC of their activities include immunosuppressive; antiarthritic;  
CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,  
CC antagonists and agonists may be useful in treating, preventing and/or  
CC diagnosing diseases and disorders such as autoimmune diseases  
CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms  
CC of the breast or liver, cerebrovascular disorders e.g. cardiac arrest,  
CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous  
CC system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
CC The polypeptides can also be used to aid wound healing and epithelial  
CC cell proliferation, to prevent skin aging due to sunburn, to maintain  
CC organs before transplantation, for supporting cell culture of primary  
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
CC also be used as a food additive or preservative to increase or decrease  
CC storage capabilities. AAC80522-C80530 and AAB45119 represent sequences  
CC used in the isolation and characterisation of the genes and proteins of  
CC the invention.  
XX Sequence 403 AA;  
SQ

Query Match  
Best Local Similarity 97.5%; Score 2121; DB 21; Length 403;  
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5 RWGPIPTSSQTNRSSFDLLPREFLVEHDPPLHQPANKPKPPTMLDIPSPCSLTI 64  
DB 1 RWGPIPTSSQTNRSSFDLLPREFLVEHDPPLHQPANKPKPPTMLDIPSPCSLTI 60  
QY 65 HTIQLIQRRLRLNLATAQNOQQTGKTESEPLPCGSPPLDILLDCNPN 124  
DB 61 HTIQLIQRRLRLNLATAQNOQQTGKTESEPLPCGSPPLDILLDCNPN 120  
QY 125 APFQIRHSDPSDFYRGKEVPVTELSWHSCROLLYQAVATTLAHAGFCANESVLETLTD 184  
DB 121 APFQIRHSDPSDFYRGKEVPVTELSWHSCROLLYQAVATTLAHAGFCANESVLETLTD 180  
QY 185 VAHEYCLAKFTLLFAVDREARLGOTPPPDVMEQVFEHVGISVLSLQKFWQHRKDYHS 244  
DB 181 VAHEYCLAKFTLLFAVDREARLGOTPPPDVMEQVFEHVGISVLSLQKFWQHRKDYHS 240

QY 245 YMLQISKOLSEYERIVNPEKATEDAKPVKIKEPVSVDITPPVSELEADLASGDQLPM 304  
DB YMLQISKOLSEYERIVNPEKATEDAKPVKIKEPVSVDITPPVSELEADLASGDQLPM 300  
QY 305 GVLGAQSERFSPNLEVEASPOASSAEVNASPLNLAHVKMEPQSEEGNVSGHVLGSDV 364  
DB 301 GVLGAQSERFSPNLEVEASPOASSAEVNASPLNLAHVKMEPQSEEGNVSGHVLGSDV 360  
QY 365 FEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORC 407  
DB 361 FEPMGSMSEAGIPQSPDDSDSSYSGSHSTDSLMGSSPVFNORC 403  
RESULT 5  
AAB45217  
ID AAB45217 standard; Protein; 404 AA.  
XX AAB45217;  
XX 12-FEB-2001 (first entry)  
XX Human secreted protein sequence encoded by gene 41 SEQ ID NO:158.  
XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; autoimmune disease; hyperproliferative disorder;  
KW cerebrovascular disorder; cerebrovascular disorder; wound healing;  
KW nervous system disorder; aging; chemotaxis.  
OS Homo sapiens.  
XX WO200058467-A1.  
XX 05-OCT-2000.  
XX 22-MAR-2000; 2000WO-US07505.  
XX 26-MAR-1999; 99US-0126502.  
XX 17-DEC-1999; 99US-0172410.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI; 2000-611712/58.  
XX N-PSDB; AAC80571.  
XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
XX Disclosure; Page 60-61; 440pp; English.  
XX Polynucleotide sequences AAC80531-C80580 represent cDNA encoding human  
CC secreted proteins AAB45120-B45169. Sequences AAB45170-B45225 represent  
CC alternative polypeptides encoded by the genes, and amino acid sequences  
CC to which they are homologous. The genes and proteins have activities  
CC dependent on the tissues and cells in which they are expressed. Examples  
CC of their activities include immunosuppressive; antiarthritic;  
CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
CC cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,  
CC antagonists and agonists may be useful in treating, preventing and/or  
CC diagnosing diseases and disorders such as autoimmune diseases  
CC e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms  
CC of the breast or liver, cerebrovascular disorders e.g. cardiac arrest,  
CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous  
CC system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
CC The polypeptides can also be used to aid wound healing and epithelial  
CC cell proliferation, to prevent skin aging due to sunburn, to maintain  
CC organs before transplantation, for supporting cell culture of primary





Db 181 QKFWHRTKDYHVMLOLQISKOLSEYERIVNPEKATEDAKPVKKEEVPVSDITFPVSEEL 240  
 QY 292 EADLASGQSLPMGVLGAQSERFSPNLEVEASPOASSAEVNASPLWNLHAHKMEPQSEEE 351  
 Db 241 EADLASGQSLPMGVLGAQSERFSPNLEVEASPOASSAEVNASPLWNLHAHKMEPQSEEE 300  
 QY 352 GNVSGHGVGLSDVFEFPMGSEAGIPQSPDSDSSYSGSHSTDSLMGSSPVFNQCKRKM 411  
 Db 301 GNVSGHGVGLSDVFEFPMGSEAGIPQSPDSDSSYSGSHSTDSLMGSSPVFNQCKRKM 360  
 QY 412 RKI 414  
 Db 361 RKI 363

RESULT 7  
 AAG98652  
 ID AAG98652 standard; Protein; 199 AA.  
 AC AAG98652;  
 XX  
 DT 21-SEP-2001 (first entry)  
 XX  
 DE Human cell death protective cDNA clone CNI-00714 ORF9 protein, SEQ:93.  
 XX  
 KW Cell death protective; apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
 KW stroke; cerebral infarction; ischaemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200145638-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 11-DEC-2000; 2000WO-US33547.  
 XX  
 PR 14-DEC-1999; 99US-0461697.  
 XX  
 PA (COGE-) COGENT NEUROSCIENCE INC.  
 XX  
 PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
 XX  
 DR WPI; 2001-390297/41.  
 XX  
 DR N-PSDB: AAH84170, AAH84179.  
 XX  
 PT Novel protective sequence polynucleotides and polypeptides, used to  
 PT identify modulators of their expression and activity, which are used in  
 PT to treat central nervous system conditions, diseases and disorders -  
 XX  
 PS Claim 1; Fig 6I; 325pp; English.  
 XX  
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84226, AAH84265,  
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
 CC while the remaining nucleic acid sequences within the range given above  
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
 CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
 CC protective ORFs. The cell death protective cDNA clones are able to  
 CC prevent, delay or reverse progression through the apoptotic or necrotic

CC pathways when injected into a cell predisposed to or undergoing cell  
 CC death. The cell death protective nucleic acids and polypeptides can be  
 CC used in the diagnosis and treatment of disorders associated with cell  
 CC death, and to screen for compounds which modulate their activity or  
 CC expression. Such modulators, preferably a small organic molecule, an  
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
 CC cell death-related diseases. Such diseases include those associated with  
 CC the central nervous system including psychiatric or neurological  
 CC disorders, especially ischaemia-related conditions such as strokes, and  
 CC also includes neurodegenerative disorders such as Alzheimer's disease,  
 CC Huntington's disease, or Parkinson's disease. The modulators may also be  
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
 CC vascular diseases such as ischaemic encephalopathy or cerebral  
 CC infarction; eye conditions such as diabetic retinopathy or macular  
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
 CC respiratory conditions such as asthma or chronic obstructive pulmonary  
 CC disease; neoplastic conditions such as cancers or benign tumours; blood  
 CC cell conditions such as anaemia; gastrointestinal conditions such as  
 CC gastritis or ulcerative colitis; liver conditions such as biliary  
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
 CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.  
 XX  
 SQ Sequence 199 AA;

Query Match 47.0%; Score 1022; DB 22; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-86;  
 Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 MEQVFHEVGVGSLVLSLOKFWQHRKTDYHVMLOLQISKOLSEYERIVNPEKATEDAKPVKI 275  
 Db 1 MEQVFHEVGVGSLVLSLOKFWQHRKTDYHVMLOLQISKOLSEYERIVNPEKATEDAKPVKI 60  
 QY 276 KEEPVSDDITFPVSEEELEADLASGQSLPMGVLGAQSERFSPNLEVEASPOASSAEVNASP 335  
 Db 61 KEEPVSDDITFPVSEEELEADLASGQSLPMGVLGAQSERFSPNLEVEASPOASSAEVNASP 120  
 QY 336 LWNLAHKMEPQSEEGNVSGHGVGLSDVFEFPMGSEAGIPQSPDSDSSYSGSHSTDS 395  
 Db 121 LWNLAHKMEPQSEEGNVSGHGVGLSDVFEFPMGSEAGIPQSPDSDSSYSGSHSTDS 180  
 QY 396 LMGSSPVFNQCKRMRKI 414  
 Db 181 LMGSSPVFNQCKRMRKI 199

## RESULT 8

AAV59941  
 ID AAV59941 standard; Protein; 198 AA.

XX AAV59941;

XX 31-JAN-2000 (first entry)

XX Human endometrium tumour EST encoded protein 1.

XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
 KW treatment; uterine; gene therapy; expressed sequence tag.

OS Homo sapiens.

XX DEL9817948-AL.

XX 21-OCT-1999.

XX 17-APR-1998; 98DE-1017948.

XX 17-APR-1998; 98DE-1017948.

XX



Db 1 MLIQISQISEYERIVNPEKATEDAKPVKIKKEPVSDITFPVSELEADLAGSQSLPMG 60  
 QY 306 VLGAQSERFPNLEVEASQASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSDVF 365  
 D 61 VLGAQSERFPNLEVEASQASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSDVF 120  
 QY 366 EEPMSGMEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 414  
 D 121 EEPMSGMEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 169

RESULT 10  
 AAG98656  
 ID AAG98656 standard; Protein; 111 AA.  
 XX  
 AC AAG98656;  
 XX  
 DT 21-SEP-2001 (first entry)  
 XX  
 DE Human cell death protective cDNA clone CNI-00714 ORF13 protein, SEQ:101.  
 XX  
 KW Cell death protective; apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischaemic encephalopathy;  
 KW stroke; cerebral infarction; ischaemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200145638-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 11-DEC-2000; 2000WO-US33547.  
 XX  
 PR 14-DEC-1999; 99US-0461697.  
 XX  
 PA (COGE-) COGENT NEUROSCIENCE INC.  
 XX  
 PI Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz IC;  
 XX  
 DR WPI; 2001-390297/41.  
 DR N-PSDB; AAH84170, AAH84183.  
 XX  
 PT Novel protective sequence polynucleotides and polypeptides, used to  
 PT identify modulators of their expression and activity, which are used in  
 PT to treat central nervous system conditions, diseases and disorders -  
 XX  
 PS Claim 1; Fig 6M; 325pp; English.  
 XX  
 CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84226, AAH84265,  
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
 CC while the remaining nucleic acid sequences within the range given above  
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
 CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
 CC protective ORFs. The cell death protective cDNA clones are able to  
 CC prevent, delay or reverse progression through the apoptotic or necrotic  
 CC pathways when injected into a cell predisposed to or undergoing cell  
 CC death. The cell death protective nucleic acids and polypeptides can be  
 CC used in the diagnosis and treatment of disorders associated with cell  
 CC death, and to screen for compounds which modulate their activity or

CC expression. Such modulators, preferably a small organic molecule, an  
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
 CC cell death-related diseases. Such diseases include those associated with  
 CC the central nervous system including psychiatric or neurological  
 CC disorders, especially ischaemia-related conditions such as strokes, and  
 CC also includes neurodegenerative disorders such as Alzheimer's disease,  
 CC Huntington's disease, or Parkinson's disease. The modulators may also be  
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
 CC vascular diseases such as ischaemic encephalopathy or cerebral  
 CC infarction; eye conditions such as diabetic retinopathy or macular  
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
 CC respiratory conditions such as asthma or chronic obstructive pulmonary  
 CC disease; neoplastic conditions such as cancers or benign tumours; blood  
 CC cell conditions such as anaemia; gastrointestinal conditions such as  
 CC gastritis or ulcerative colitis; liver conditions such as biliary  
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
 CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.  
 XX  
 SQ Sequence 111 AA;  
 Query Match 26.3%; Score 572; DB 22; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1e-44;  
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 304 MGVLGAQSERFPNLEVEASQASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSD 363  
 D 1 MGVLGAQSERFPNLEVEASQASSAEVNASPLNLAHVKNMPEQSEEGNVSGHVLGSD 60  
 QY 364 VFEPMSGMEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 414  
 D 61 VFEPMSGMEAGIPQSPDDSDSYGSHSTDSLMGSSPVFNQCKKMRKI 111

RESULT 11  
 AAG98658  
 ID AAG98658 standard; Protein; 71 AA.  
 XX  
 AC AAG98658;  
 XX  
 DT 21-SEP-2001 (first entry)  
 XX  
 DE Human cell death protective cDNA clone CNI-00714 ORF15 protein, SEQ:105.  
 XX  
 KW Cell death protective; apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
 KW stroke; cerebral infarction; ischaemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200145638-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 11-DEC-2000; 2000WO-US33547.  
 XX  
 PR 14-DEC-1999; 99US-0461697.

XX PA (COGE-) COGENT NEUROSCIENCE INC.  
XX KW Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
XX KW WPI; 2001-390297/41.  
XX DR N-PSDB; AAH84170, AAH84185.  
XX KW Novel protective sequence polynucleotides and polypeptides, used to  
XX PT identify modulators of their expression and activity, which are used in  
XX PT to treat central nervous system conditions, diseases and disorders -  
XX PS Claim 1; Fig 60; 325pp; English.  
XX KW Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
XX CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
XX CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
XX CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
XX CC while the remaining nucleic acid sequences within the range given above  
XX CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
XX CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
XX CC protective ORFs. The cell death protective cDNA clones are able to  
XX CC prevent, delay or reverse progression through the apoptotic or necrotic  
XX CC pathways when injected into a cell predisposed to or undergoing cell  
XX CC death. The cell death protective nucleic acids and polypeptides can be  
XX CC used in the diagnosis and treatment of disorders associated with cell  
XX CC death, and to screen for compounds which modulate their activity or  
XX CC expression. Such modulators, preferably a small organic molecule, an  
XX CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
XX CC cell death-related diseases. Such diseases include those associated with  
XX CC the central nervous system including psychiatric or neurological  
XX CC disorders, especially ischaemia-related conditions such as strokes, and  
XX CC also includes neurodegenerative disorders such as Alzheimer's disease,  
XX CC Huntington's disease, or Parkinson's disease. The modulators may also be  
XX CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
XX CC vascular diseases such as ischaemic encephalopathy or cerebral  
XX CC infarction; eye conditions such as diabetic retinopathy or macular  
XX CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
XX CC respiratory conditions such as asthma or chronic obstructive pulmonary  
XX CC disease; neoplastic conditions such as cancers or benign tumours; blood  
XX CC cell conditions such as anaemia; gastrointestinal conditions such as  
XX CC gastritis or ulcerative colitis; liver conditions such as biliary  
XX CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
XX CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
XX CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
XX CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
XX CC nucleic acids may additionally be used to generate animal models of  
XX CC cell death-associated disorders. The present sequence represents a  
XX CC cell death protective polypeptide.  
XX SQ Sequence 71 AA;  
Query Match 17.1%; Score 371; DB 22; Length 71;  
Best Local Similarity 100.0%; Pred. No. 1.7e-26;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 344 MEPQSEEGNVSGHVGIVSDVEEPMGMSGAGIPQSDSSYGSHTDSLGMSSPVF 403  
Db 1 MEPQSEEGNVSGHVGIVSDVEEPMGMSGAGIPQSDSSYGSHTDSLGMSSPVF 60  
QY 404 NORCKRMRKI 414  
Db 61 NORCKRMRKI 71  
RESULT 12  
AAG98662  
ID AAG98662 standard; Protein; 46 AA.  
XX AC AAG98662;  
XX DT 21-SEP-2001 (first entry)  
XX XX

DE XX Human cell death protective cDNA clone CN1-00714 ORF19 protein, SEQ:113.  
XX KW Cell death protective; apoptosis; necrosis; human; drug screening;  
XX KW cell death-associated disorder; neurological disorder; ischaemia-related disorder;  
XX KW psychiatric disorder; neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
XX KW stroke; cerebral infarction; ischaemic encephalopathy;  
XX KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
XX KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
XX KW vascular degeneration; ophthalmological disorder; diabetic retinopathy;  
XX KW macular degeneration; hypertension; myocardial infarction;  
XX KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
XX KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
XX KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
XX KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
XX KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
XX KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
XX KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.  
XX OS Homo sapiens.  
XX PN WO200145638-A2.  
XX XX 28-JUN-2001.  
XX PD 11-DEC-2000; 2000WO-US33547.  
XX PF 14-DEC-1999; 99US-0461697.  
XX XX (COGE-) COGENT NEUROSCIENCE INC.  
XX PA Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
XX PI WPI; 2001-390297/41.  
XX XX N-PSDB; AAH84170, AAH84189.  
XX DR Novel protective sequence polynucleotides and polypeptides, used to  
XX XX identify modulators of their expression and activity, which are used in  
XX PT to treat central nervous system conditions, diseases and disorders -  
XX PS Claim 1; Fig 65; 325pp; English.  
XX KW Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
XX CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
XX CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84210, AAH84226, AAH84265,  
XX CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
XX CC while the remaining nucleic acid sequences within the range given above  
XX CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
XX CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
XX CC protective ORFs. The cell death protective cDNA clones are able to  
XX CC prevent, delay or reverse progression through the apoptotic or necrotic  
XX CC pathways when injected into a cell predisposed to or undergoing cell  
XX CC death. The cell death protective nucleic acids and polypeptides can be  
XX CC used in the diagnosis and treatment of disorders associated with cell  
XX CC death, and to screen for compounds which modulate their activity or  
XX CC expression. Such modulators, preferably a small organic molecule, an  
XX CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
XX CC cell death-related diseases. Such diseases include those associated with  
XX CC the central nervous system including psychiatric or neurological  
XX CC disorders, especially ischaemia-related conditions such as strokes, and  
XX CC also includes neurodegenerative disorders such as Alzheimer's disease,  
XX CC Huntington's disease, or Parkinson's disease. The modulators may also be  
XX CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
XX CC vascular diseases such as ischaemic encephalopathy or cerebral  
XX CC infarction; eye conditions such as diabetic retinopathy or macular  
XX CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
XX CC respiratory conditions such as asthma or chronic obstructive pulmonary  
XX CC disease; neoplastic conditions such as cancers or benign tumours; blood  
XX CC cell conditions such as anaemia; gastrointestinal conditions such as  
XX CC gastritis or ulcerative colitis; liver conditions such as biliary  
XX CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
XX CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
XX CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
XX CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
XX CC nucleic acids may additionally be used to generate animal models of  
XX CC cell death-associated disorders. The present sequence represents a  
XX CC cell death protective polypeptide.

CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.

XX  
 SQ Sequence 46 AA;

Query Match 11.0%; Score 239; DB 22; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-14;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 MSGMEAGIPOSDDSSSYGSHSTDSLMGSSPVFNQCKRMKRI 414  
 |||||  
 DB 1 MSGMEAGIPOSDDSSSYGSHSTDSLMGSSPVFNQCKRMKRI 46

RESULT 13  
 AAG98663  
 ID AAG98663 standard; Protein; 43 AA.  
 AC AAG98663;  
 DT 21-SEP-2001 (first entry)  
 XX Human cell death protective cDNA clone CNI-00714 ORF20 protein, SEQ:115.

CC Cell death protective; apoptosis; necrosis; human; drug screening;  
 KW cell death-associated disorder; central nervous system disorder;  
 KW psychiatric disorder; neurological disorder; ischaemia-related disorder;  
 KW stroke; cerebral infarction; ischaemic encephalopathy;  
 KW neurodegenerative disorder; Alzheimer's disease; Huntington's disease;  
 KW Parkinson's disease; infection; meningitis; malaria; trypanosomiasis;  
 KW vascular disease; ophthalmological disorder; diabetic retinopathy;  
 KW macular degeneration; hypertension; myocardial infarction;  
 KW atherosclerosis; respiratory disorder; asthma; transgenic animal;  
 KW chronic obstructive pulmonary disease; neoplastic condition; cancer;  
 KW benign tumour; anaemia; gastrointestinal disorder; gastritis;  
 KW ulcerative colitis; liver disease; biliary cirrhosis; kidney disorder;  
 KW glomerulonephritis; cystitis; endometriosis; endocrine disorder;  
 KW Grave's disease; Hashimoto's thyroiditis; skin condition; dermatitis;  
 KW urticaria; immune disorder; acquired immunodeficiency syndrome; AIDS.

XX Homo sapiens.  
 OS  
 XX WO200145638-A2.  
 PN  
 XX 28-JUN-2001.  
 PD  
 XX 11-DEC-2000; 2000WO-US33547.  
 PF  
 XX 14-DEC-1999; 99US-0461697.  
 PR  
 XX (COGE-) COGENT NEUROSCIENCE INC.  
 PA  
 XX Lo DC, Barney S, Thomas MB, Portbury SD, Puranam K, Katz LC;  
 PI  
 XX WPI; 2001-390297/41.  
 XX N-PSDB; AAH84170, AAH84190.  
 DR  
 XX Novel protective sequence polynucleotides and polypeptides, used to  
 PT identify modulators of their expression and activity, which are used in  
 PT to treat central nervous system conditions, diseases and disorders -  
 XX  
 PS Claim 1; Fig 6T; 325pp; English.

CC Sequences AAH84132-AAH84370 represent human nucleic acid sequences which  
 CC protect against cell death (i.e., apoptosis or necrosis). Sequences  
 CC AAH84132, AAH84145, AAH84170, AAH84201, AAH84226, AAH84265,  
 CC AAH84281, AAH84315 and AAH84367 represent 10 full-length cDNA clones,  
 CC while the remaining nucleic acid sequences within the range given above  
 CC represent the open reading frames (ORFs) of these cDNA clones. Sequences  
 CC AAG98610-AAG98829 represent the polypeptides encoded by the cell death  
 CC protective ORFs. The cell death protective cDNA clones are able to  
 CC prevent, delay or reverse progression through the apoptotic or necrotic

CC pathways when injected into a cell predisposed to or undergoing cell  
 CC death. The cell death protective nucleic acids and polypeptides can be  
 CC used in the diagnosis and treatment of disorders associated with cell  
 CC death, and to screen for compounds which modulate their activity or  
 CC expression. Such modulators, preferably a small organic molecule, an  
 CC antibody, a ribozyme, or an antisense molecule, can also be used to treat  
 CC cell death-related diseases. Such diseases include those associated with  
 CC the central nervous system including psychiatric or neurological  
 CC disorders, especially ischaemia-related conditions such as strokes, and  
 CC also includes neurodegenerative disorders such as Alzheimer's disease,  
 CC Huntington's disease, or Parkinson's disease. The modulators may also be  
 CC used to treat infections such as meningitis, malaria, or trypanosomiasis;  
 CC vascular diseases such as ischaemic encephalopathy or cerebral  
 CC infarction; eye conditions such as diabetic retinopathy or macular  
 CC degeneration; hypertension; myocardial infarction; atherosclerosis;  
 CC respiratory conditions such as asthma or chronic obstructive pulmonary  
 CC disease; neoplastic conditions such as cancers or benign tumours; blood  
 CC cell conditions such as anaemia; gastrointestinal conditions such as  
 CC gastritis or ulcerative colitis; liver conditions such as biliary  
 CC cirrhosis; kidney disorders such as glomerulonephritis; cystitis;  
 CC endometriosis; endocrine disorders such as Grave's disease or Hashimoto's  
 CC thyroiditis; skin conditions such as dermatitis or urticaria; or immune  
 CC system disorders such as acquired immunodeficiency syndrome (AIDS). The  
 CC nucleic acids may additionally be used to generate animal models of  
 CC cell death-associated disorders. The present sequence represents a  
 CC cell death protective polypeptide.

XX  
 SQ Sequence 43 AA;

Query Match 10.3%; Score 224; DB 22; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-13;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 MSEAGIPOSDDSSSYGSHSTDSLMGSSPVFNQCKRMKRI 414  
 |||||  
 DB 1 MSEAGIPOSDDSSSYGSHSTDSLMGSSPVFNQCKRMKRI 43

RESULT 14  
 ABB62430  
 ID ABB62430 standard; Protein; 359 AA.  
 XX  
 AC ABB62430;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 14082.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PD  
 XX 27-SEP-2001.  
 PF  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR  
 XX 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 XX N-PSDB; ABL06533.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX



**THIS PAGE BLANK (USPTO)**





# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

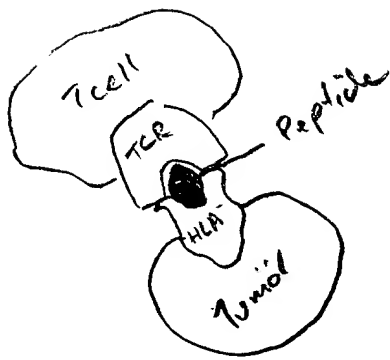
➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed form to STIC/Biotech-Chem Library CM1 – Circ. Desk





---

Tumor Ag From Squamous cell carcinoma

HLA-A24 or HLA-A26